Electrical Power and Energy Systems 132 (2021) 107176

Contents lists available at ScienceDirect

International Journal of Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

A taxonomical review on recent artificial intelligence applications to PV integration into power grids

Cong Feng [a], Yuanzhi Liu [a], Jie Zhang [a,*]

[a] Department of Mechanical Engineering, The University of Texas at Dallas, Richardson, TX 75080, USA

ARTICLE INFO

Keywords:
Solar forecasting
Solar array and fault detection
Optimization
Solar optimal control
Text mining review

ABSTRACT

The exponential growth of solar power has been witnessed in the past decade and is projected by the ambitious policy targets. Nevertheless, the proliferation of solar energy poses challenges to power system operations, mostly due to its uncertainty, locational specificity, and variability. The prevalence of smart grids enables artificial intelligence (AI) techniques to mitigate solar integration problems with massive amounts of solar energy data. Different AI subfields (e.g., machine learning, deep learning, ensemble learning, and metaheuristic learning) have brought breakthroughs in solar energy, especially in its grid integration. However, AI research in solar integration is still at the preliminary stage, and is lagging behind the AI mainstream. Aiming to inspire deep AI involvement in the solar energy domain, this paper presents a taxonomical overview of AI applications in solar photovoltaic (PV) systems. Text mining techniques are first used as an assistive tool to collect, analyze, and categorize a large volume of literature in this field. Then, based on the constructed literature infrastructure, recent advancements in AI applications to solar forecasting, PV array detection, PV system fault detection, design optimization, and maximum power point tracking control problems are comprehensively reviewed. Current challenges and future trends of AI applications in solar integration are also discussed for each application theme.

1. Introduction

Energy demand, proportional to the living population and economic development, has been increasing exponentially since the 21st century [1]. To prevent the energy crisis while protecting the ecosystem from fossil fuel-generated pollution, renewable energy is endowed with more responsibilities in sustaining the energy demand. Although the levelized cost of electricity (LCOE) of utility-scale photovoltaic (PV) has fallen by more than 77% in the last decade (i.e., from 0.370 $/kWh in 2010 to 0.085 $/kWh in 2018), solar PV is still not cost-competitive compared to other energy commodities in the wholesale market (e.g., LCOE of onshore wind is 0.055 $/kWh) [2]. Therefore, a broad range of policy incentives, including fiscal, regulatory, and marketing instruments, have been introduced by governments to promote solar penetrations [3]. For example, feed-in-tariffs play a dominant role in Germany and Spain. The U.S. relies more on renewable energy portfolio standards, federal tax credits, and renewable energy certificates. The fixed budget incentive program (e.g., subsidies) determines the solar installation scale in China. As a result, solar energy, mostly solar PV energy[1], has experienced phenomenal growth. Specifically, the global solar PV capacity reached 480 GW and another 137.5 GW PV was installed in 2019 throughout the world. Fig. 1 shows the top 10 countries based on PV capacity. Most PV installations were added in the last few years, where China contributed more than half of the addition.

As one of the largest man-made systems, the power and energy system needs to be operated with high stability and low operational costs. Hence, the integration of uncertain and variable solar energy into power systems is complicated and challenging. A large collection of research projects are supported by governments and industries to mitigate these challenges, among which artificial intelligence (AI) is one of the emerging topics. In these projects, AI techniques are applied to different stages of solar integration, such as resource assessment, design optimization, optimal control, system monitoring, generation or irradiance forecasting, etc. For example, the U.S. Department of Energy (DOE) funded two rounds of projects, called Solar Forecasting and Solar Forecasting 2, in 2012 and 2017, respectively, to improve the solar forecasting, where AI plays a pivotal role. These projects are expected to improve the management of solar power's variability and uncertainty, enabling its more reliable and cost-effective integration into the grid [4,5]. In addition, the DOE Advanced Research Projects Agency–Energy (ARPA-E) supported a projects to promote the AI applications in the

* Corresponding author.
E-mail address: jiezhang@utdallas.edu (J. Zhang).

[1] This paper focus on solar PV systems, "solar energy" refers to energy provided by solar PV systems in this paper.

https://doi.org/10.1016/j.ijepes.2021.107176
Received 26 September 2020; Received in revised form 1 March 2021; Accepted 30 April 2021

Nomenclature

ACO	Ant colony optimization
AI	Artificial intelligence
ANN	Artificial neural network
CNN	Convolutional neural network
DNI	Direct normal irradiance
DNN	Deep neural network
DT	Decision tree
ELM	Extreme learning machine
FD	Fault detection
FFNN	Feedforward neural network
FLC	Fuzzy logic control
GA	Genetic algorithm
GBM	Gradient boosting machine
GHI	Global horizontal irradiance
GNI	Global normal irradiance
GP	Gaussian process
kNN	K-nearest neighbors
LDA	Latent Dirichlet allocation
LOEE	Loss of energy expectation
LOLP	Loss of load probability
LPSP	Loss of power supply probability
LSTM	Long short-term memory
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MBE	Mean bias error
ML	Machine learning
MLP	Multi-layer perceptron
MPPT	Maximum power point tracking
NWP	Numerical weather prediction
PSO	Particle swarm optimization
PV	Photovoltaic
RBF	Radial basis function
RBFNN	Radial basis function neural network
RBR	Red blue ratio
ResNet	Residual network
RF	Random forest
RMSE	Root mean square error
ROC	Receiver operating characteristic
SA	Simulated annealing
SS	Skill score
SVM	Support vector machine
SVR	Support vector regression
TAC	Total annual cost
TS	Tabu search
TSC	Total system cost

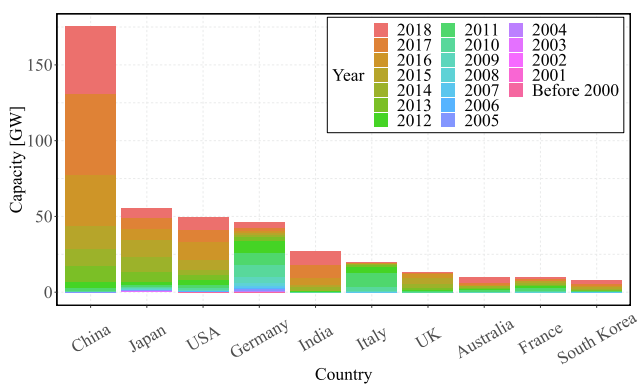

Fig. 1. The top ten countries based on PV installation (by 2018).

energy domain, where solar energy is one of the major topics [6]. The U. S. National Science Foundation (NSF) supports the Energy, Power, Control, and Networks Program, where AI applications to solar energy is an emerging topic [7]. The awarded projects covered machine learning (ML)-based PV fault detection [8], PV array connected topology optimization [9], and PV voltage regulation [10]. China is also promoting AI research in solar energy by financing projects through various programs and foundations, such as the National Key R&D Program of China, the National Natural Science Foundation of China, Fundamental Research Funds for the Central Universities, and the National Key Research and Development Program of China. Meanwhile, a large number of projects are granted by the European Union through programs, such as the Horizon 2020 research and innovation program and European Regional Development Fund, to support AI-based solar energy research.

In addition to academic research, the power and energy industry has also been driving applications of AI techniques to power and energy system operations in the past decades. For example, ML techniques helped IBM improve solar forecast accuracy by 30% from various data sources and provide forecasting services for several utilities. Moreover, ML-enhanced probabilistic renewable energy forecasting products were developed for three German transmission system operators (TSOs) in the EWeLiNE and Gridcast projects [11]. The GridSense, an intelligent distribution system management product developed by Alpiq in Switzerland, proactively and automatically controls the loads and storage by learning consumer behaviour, system parameters, and weather conditions with AI [12]. Ampacimon developed an AI-based self-learning meteorological network for weather-dependent overhead line operation, called PrognoNetz [13]. In power systems with high solar penetrations, AI techniques are used to assist the electricity market design and operation. For example, an AI-based electricity price forecasting algorithm, i.e., the Pan-European Hybrid Electricity Market Integration Algorithm, was developed to forecast day-ahead electricity prices across Europe and allocate cross-border transmission capacity for twenty-five European countries [14]. Several AI-enabled robots were developed for PV farm operations and maintenance, such as PV panel cleaning [15] and PV farm monitoring [16].

Large attention has been given toward solar AI topics, therefore, leading to abundant literature. More than 35,700 and 44,500 results turned up when searching "solar photovoltaic AND artificial intelligence" and "solar photovoltaic AND machine learning" in Google Scholar (by 2020–01-01), respectively. A collection of journals encourage publications on this topic by organizing special issues, accepting new types of articles, or even highlighting solar AI research in their publishing scope. For example, IEEE Transactions on Industrial Informatics opened a special section on developments in AI for industrial informatics to discover intelligent decision-making solutions for smart cities, smart grids, and smart homes [17]. This special section received more than 100 papers, among which AI applications to solar PV systems is a main topic. IEEE Transactions on Intelligent Systems proposed a special issue in AI in power systems and energy markets to explore the AI benefits to power systems with distributed generations, such as solar PV [18]. In addition, Energies summarized AI-based solar energy application cases in several special issues, such as energy harvesting [19]. Moreover, Solar Energy and Journal of Renewable and Sustainable Energy encourage authors to submit new types of articles, such as data articles, to promote AI research in solar [20,21]. Journal of Renewable and Sustainable Energy explicitly indicated that renewable energy resource assessment, analysis, and forecasting, where AI-related research thrives, are primary

fields of interest to the journal readership [21].

Solar AI research usually requires inter-disciplinary knowledge from various domains, such as power engineering, energy science, meteorology, data science, computer science, economics, etc. This interdisciplinary nature in conjunction with the large body of literature makes the task of identifying relevant studies in an unbiased way for inclusion in systematic reviews both complex and time consuming [22]. Although the review papers in this field have provided excellent reviews on solar AI research, their comprehensiveness might not be optimal. First, the papers covered by the review papers were relatively limited. Most review papers only summarized around 100 papers, which is far from comprehensive, compared to the large number of publications. Second, it is not clear what criteria being used for authors to select, categorize, and summarize papers in the review [23]. Largely relying on searching results from the bibliographic databases, authors were hardly exposed to every paper with an equal chance. Therefore, there exist biases in searching proper papers in academic literature databases, which in the end affects the objectivity of the literature review. Third, these review papers reviewed only one topic of solar AI research. Fourth, the text mining-based bibliometric analysis is especially beneficial to junior scholars who need an overall picture of the field, conference organizers who desire to offer mini-tracks and workshops with emerging topics, and journal editors who want to document the history and develop particular streams of research [24]. In recent years, text mining was utilized to help review literature and construct knowledge infrastructures in various domains, such as business intelligence [25], information science [24], and biomedical science [26,27]. Recently, [23] introduced text mining into the solar forecasting area and exemplified how to use text mining as an assistive approach for a systematic and comprehensive literature review. In this paper, we utilize text mining techniques as an assistive approach to collect, analyze, and categorize the bibliography of the recent 10 years' AI-based solar energy research. Then, based on the constructed bibliographical dictionary, a taxonomical review of four main application themes is performed, which are solar forecasting, PV array and system fault detection, PV system design optimization, and control.

The remainder of the paper is organized as follows. The text mining-based bibliographic analysis and categorization are described in Section 2. Section 3 reviews ML, ensemble learning, and deep learning methods in solar forecasting. AI-based PV array detection and PV fault detection are reviewed in Section 4. Section 5 reviews ML and metaheuristic learning applications in PV system design optimization. Optimal control methods that rely on AI techniques are reviewed in Section 6. Section 7 reviews papers on other relevant AI-based solar topics. Section 9 summarizes and concludes the paper.

2. Text mining

Text mining, a subset of data mining, derives patterns and trends from raw textual data through statistical learning approaches. Compared to the traditional literature review that is likely exposed to a limited volume of published work, text mining is able to screen almost the entire literature records of a specific discipline. The uniqueness of text mining-assisted literature review is advanced by the easy accessibility to several bibliographic databases, such as Google Scholar.

2.1. Text mining process

A successful literature review with text mining techniques requires six major steps [28,23]. First, problems and goals should be well-defined. In this research, we seek to answer the following questions by text mining the literature databases:

Q1. Which journals should we refer to in solar AI research?
Q2. What are the most important and frequent terms in this area?
Q3. What are the hot topics in solar AI research and how can we categorize solar AI research?
Q4. Which papers should we review for each solar AI topic?

To answer those questions, in the second step, proper textual data should be identified and collected. We rely on three of the most popular bibliographic databases, i.e., Google Scholar, Web of Science, and Scopus, to mitigate the limitations of each single database [29]. The searching query is: Topic=("solar photovoltaic" AND ("artificial intelligence" OR "machine learning")) AND (Year Published = 2010–2020). The search is conducted in every database year by year to avoid database exporting limitations (e.g., Google Scholar only returns the 1,000 most cited papers). The total numbers of results retrieved from Google Scholar, Web of Science, and Scopus are, respectively 2,516, 1,433, and 1,006 (by 2019–12-01). Then, all the results are combined with duplications removed, which leads to a total of 3,420 bibliographic records. However, this textual data might still be problematic. For example, some non-scholarly (e.g., handbooks, newsletters, and course notes) or irrelevant records (e.g., astronomy, material science, and agriculture) are contained in the data. To refine the results, a collection of keywords are used to filter out improper results by checking the title, source, and type, which are listed in Table 1. The quality control reduces the paper number to 2,772.

In the third step, the textual data is preprocessed and organized for later analysis. This step is realized by applying a sequence of techniques, including creating the corpus, tokenization, upper-to-lower case conversion, stop-word removal, whitespace removal, punctuation removal, stemming, document-term matrix generation. Details of these techniques can be found in [28]. Two packages, namely, tm and corpus, in **R** are used for this purpose. With the document-term matrix, the rest three steps extract features, analyze patterns, and draw conclusions to identify relevant review papers, top journals, categorize technical themes, and create the literature infrastructure for further review.

2.2. Bibliometric analysis and taxonomy

Fig. 2 shows the number of relevant publications in the last ten years, which indicates an increasing trend (note that publications in the last month of 2019 are not available by the time of online search). Before 2015, AI is not a mainstream technique in the solar PV energy toolbox. Most research in this period applied relatively simple ML algorithms, such as fuzzy logic, in the solar PV system design [30], control [31], and forecasting [32]. The literature body of solar AI research grows rapidly from 2015 to 2019. Considering the typical publication cycle, the two waves of solar AI publication increase are possibly due to the prevalence of ML and deep learning in the early 2010s and years after 2015. More advanced AI techniques bring not only quantitative increase but also breakthroughs in the existing research, such as image-based forecasting [33] and reinforcement learning-based microgrid control [34]. New solar research themes also emerge along with the new techniques in AI, such as PV array detection [35].

As part of the technology infrastructure, it is also of interest to identify the journals in this topic. Fig. 3 ranks journals by the number of

Table 1
Terms used to remove or select proper records.

Source (without)	Title (without)	Type (with)
Conference	Biogeochemical	Article
Workshop	Chemical	Conference paper
Coronal	Astronomical	Review
Astrophysical	Astrophysical	Chapter
Astronomy	Astronomy	Proceedings
River	Geophysical	Book
Temperature	Agricultural	
Thermoeconomic	Earth	
Forest	Space	

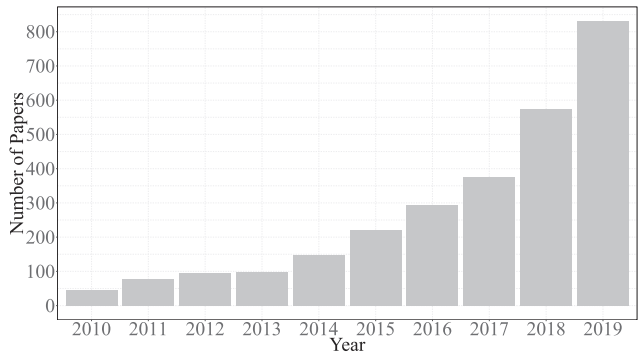

Fig. 2. The publication number by year.

solar AI papers published in the last decade. Solar Energy was leading among all other journals, with a total of 403 papers. Energies and Renewable Energy also published abundant papers. Considering the longer review process, solar AI research is also a hot topic in IEEE transactions. Different journals might have different focuses on solar AI research. For example, Energy and Buildings emphasizes in PV systems in smart home/buildings while IEEE Transactions on Power Systems focuses more on utility-scale PV integrations to transmission or distribution systems. Among the 2,772 publications, 96 of them are review papers. Nearly half of review papers (39/96) in this domain are published in Renewable and Sustainable Energy Reviews, followed by Energy Conversion and Management (4/96) and Solar Energy (4/96). While most review papers considered AI techniques as one group of methods, 20 papers directly reviewed AI applications to solar PV or renewable energy systems, which are listed in Table 2. It is observed from these review papers that forecasting, fault detection, control, and design optimization [48] are prevailing AI application cases for solar PV or hybrid renewable energy systems. Artificial neural network (ANN) and support vector machine (SVM) [38] are the two most widely-used AI algorithms and deep learning [40] is attracting high attention in recent solar research.

Another way to identify the most popular research themes or techniques is by calculating term frequencies. The term frequency is one of the key metrics for text mining, which reflects on how salient or important a term is. A term refers to a token of one word or a sequence of words, called an n-gram. Generally, a term with fewer tokens appears more frequently in the corpus. Therefore, in this research, n-grams with tokens of up to four words (i.e., unigrams, bigrams, trigrams, and teragrams) are analyzed. Fig. 4 shows the top 40 n-grams in the 2,772 titles. It is found that forecasting, optimization, control, and detection are the top four themes. Regarding the AI techniques, ANN, SVM, extreme learning machine, reinforcement learning, and particle swarm optimization are most frequently-used. The PV system is always included in hybrid renewable energy systems or home energy systems to coordinate with other types of renewable energies or electric components in this research.

Even though the primary themes can be identified by the term frequency, a more automatic taxonomy is required to group the solar AI research by categorizing publications (i.e., an unsupervised learning problem). To achieve this, a topic modeling approach, i.e., latent Dirichlet allocation (LDA), is used to discover the latent themes from the publication titles. LDA is a generative statistical model that assigns a document (i.e., a publication title in this study) to a mixture of themes based on the contributions of tokens to each theme. However, the number of themes is needed in order to assign each paper to a theme. Then, the optimal number of themes is defined by evaluating the clustering results of different numbers. We follow the same procedure as in [23], which uses four indicators [56–59] to assess the results, as shown in Fig. 5. All the four metrics range between 0 and 1. Two metrics indicate the best number of themes with the maximum, while the other two metrics with the minimal value indicate the best number of themes. In solar AI research, the optimal number of themes is five by considering the four metrics equally. These five themes' keywords are listed in Table 3. The first four themes are easily recognized from the keywords, which are forecasting, control, design optimization, and detection, which are consistent with what we observed from review papers and

Table 2
Review papers that focus on solar, renewable, or power system AI research.

Review	Key words	Journal	#Refs.
[36]	Solar forecasting	Renew Energ	105
[37]	Solar forecasting, neural network	J Clean Prod	54
[38]	Solar forecasting, support vector machine	J Clean Prod	92
[39]	Solar forecasting	IEEE Lat AM Trans	143
[40]	Solar forecasting, deep learning	Energ Convers Manage	169
[41]	Solar forecasting, neural network	Appl Sci	109
[42]	Solar forecasting, ML	IET Renew Power Gen	147
[43]	PV, control	Renew Sus Energ Rev	139
[44]	PV-wind control, neural network	Renew Sus Energ Rev	91
[45]	PV design, control	Renew Sus Energ Rev	106
[46]	PV, control	Energy	154
[47]	PV control, neural network	IET Renew Power Gen	51
[48]	Renewable energy, design optimization	Renew Sus Energ Rev	145
[49]	PV monitoring, fault detection	Renew Sus Energ Rev	95
[50]	PV system, fault detection & diagnosis	Renew Sus Energ Rev	141
[51]	PV modeling, fault detection	Renew Sus Energ Rev	156
[52]	Renewable energy, neural network	Renew Sus Energ Rev	222
[53]	Solar energy modeling, neural network	Sol Energy	161
[54]	Energy systems	Energies	114
[55]	PV integration, power quality	Renew Sus Energ Rev	94

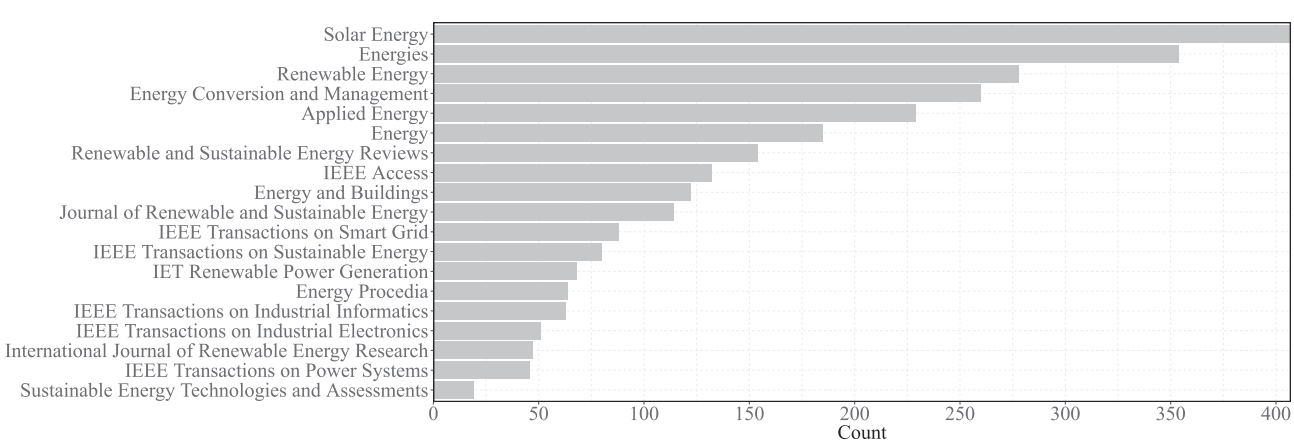

Fig. 3. The popular journals that publish solar AI research.

Fig. 4. Top 40 n-grams.

term frequencies. Specifically, solar irradiance or power forecasting is one of the most effective and efficient approaches to mitigate the solar power uncertainties in power systems and is a dominant theme in the entire solar AI research. Solar PV detection mainly includes PV array detection and fault detection. Intelligent controllers play a significant role in tracking maximum power from PV panel arrays and regulating the inverter AC power output and frequency. AI techniques also improve design optimization of PV systems, such as optimal siting and sizing in the PV to the grid, hybrid renewable system, microgrid, and PV to electrical vehicle research. Nevertheless, keywords in the last theme are vague, such as “regression”, “classification”, and “sensor”, which could be used in any solar AI research. Therefore, all the research that is not included in the previous four themes are assigned to the last theme, such as cloud tracking, weather classification, and solar datasets. Based on probabilities associated with the theme assignments, each paper is assigned to one of the five themes. The extensive review in the following sections will focus on the first four main themes.

Although text mining techniques can screen a large volume of papers and infer some information through linguistic analysis, it is only an assistive approach, which cannot replace the conventional human-centric literature reviews. A review paper typically summarizes and analyzes a few hundreds of papers, as listed in Table 2. It is almost impossible to review all the 2,772 papers collected in the text mining. Therefore, in the rest of this paper, publications that have higher assignment probabilities, annual citations, and were published more recently will be given more attention.

3. Forecasting

Integrating high penetration of solar poses challenges to reliable and economic power system operations, since the uncertain and variable solar power intensifies the frequency fluctuation, voltage instability, and harmonics. For example, cloud transient effects may intensify the power swing issues in systems with high solar penetrations [60,61]. Consequently, more reserves or energy storage are required to ensure the power system reliability [62,63]. And these issues turn to a result of

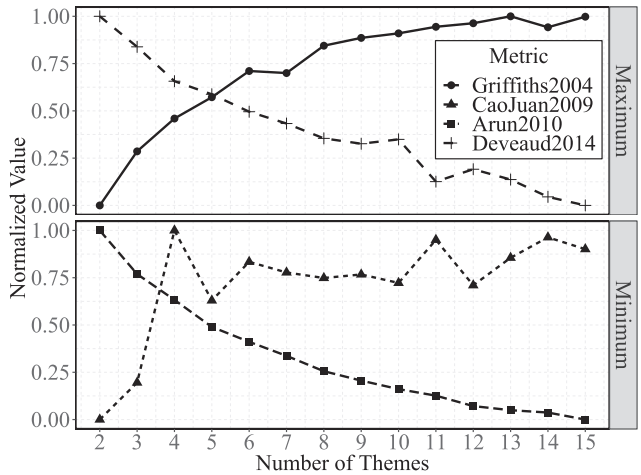

Fig. 5. Evaluation of the best number of LDA themes. The right axis indicates weather the metric should be maximized or minimized.

Table 3
Top 10 words contributing to each solar AI research theme.

Theme 1 (forecasting)	Theme 2 (detection)	Theme 3 (control)	Theme 4 (design optimization)	Theme 5 (others)
forecast	network	power	hybrid	data
irradiance	neural	system	optimization	regression
global	detection	maximum	design	estimation
short-term	analysis	track	renewable	classification
weather	fault	fuzzy	management	sensor
ensemble	plant	mppt	smart	feature
comparison	image	controller	grid	selection
evaluation	diagnosis	monitor	microgrid	dynamic
meteorological	distribution	inverter	storage	cloud
production	cluster	dynamic	distribution	remote

large solar curtailments and limiting further penetration of solar energy [64,65]. Hence, accurate solar forecasting, including both solar power and solar irradiance forecasting, becomes necessary in daily power system operations [66,67]. For instance, the California Independent System Operator needs hourly 30.5-h-ahead (30.5HA) solar forecasts in the day-ahead market and 5-min 1HA solar forecasts in the real-time market for the unit commitment and economic dispatch [68]. The Electric Reliability Council of Texas uses solar forecasts, along with load and wind forecasts, to determine daily regulation, responsive (i.e., spinning), and non-spinning reserve requirements. A detailed literature review on solar forecasting economics can be found in [69].

There are generally two categories of solar forecasting methods. The first group is physical methods, which are mainly numerical weather prediction (NWP) models. The second group is data-driven methods. Compared to physical models that use physical laws and meteorological conditions, data-driven models are generally suitable for up to 6HA forecasting. Sometimes, data-driven methods are used in post-processing to enhance the physical model, where the inputs to data-driven models are physical model outputs. We also consider this group of methods as data-driven methods in this paper. Additionally, only the advanced AI techniques are reviewed in this paper, which means traditional statistical models, such as auto-regressive moving average (ARMA) [70] models and auto-regressive moving average with exogenous variables (ARMAX) [71] models are excluded. Among data-driven models, data and algorithm are the two most important factors that influence the forecast accuracy. Hence, this section reviews AI-based forecasting methods by the input and algorithm.

Forecasting results are generally evaluated by (i) forecasting error metrics, such as mean bias error (MBE), root mean square error (RMSE), mean absolute error (MAE), (ii) error distributions or their statistics, such as forecast-observation joint distribution, Kolmogorov Smirnov Integral, skewness, kurtosis, and entropy of error distributions, (iii) skill scores (SS), and (iv) uncertainty metrics, such as pinball loss, reliability, and sharpness. Other types of metrics, such as metrics that quantify ramp characteristics and economic metrics are also used in the literature. Definitions and formulations of these standard forecast accuracy assessment metrics can be found in [72,73].

3.1. Forecasting inputs

There are two categories of approaches, according to the input source for solar forecasting [69]: models with endogenous inputs (historical forecasting target variable) and models that use exogenous inputs together with historical forecasting target variable, such as meteorological measurements, sky images, satellite data, NWP, and spatially or temporally hierarchical information. Although the response time of models that only use endogenous data may be faster [74], they cannot compete with models including exogenous data in virtually all cases.

3.1.1. Endogenous data

Endogenous data is a major relevant input to AI models, from which forecasts can be generated with satisfactory accuracy. [75] compared three forecasting models (i.e., recursive autoregressive and moving average (RARMA), coupled autoregressive and dynamical system (CARDS), and ANN) with only the past irradiance data as inputs in the 1HA–6HA global horizontal irradiance (GHI) forecasting. Similarly, only the past three GHI data points were used to train ANN models for 5-min-ahead (5MA) forecasting with optimal performance in different months and weather conditions [76]. It was found that ARMA and ANN had similar results in terms of the forecasting MBE, RMSE, MAE, and error distributions. In another solar PV power forecasting paper, [77] found that by only using solar power in the past 13 h, ANN outperformed auto regressive integrated moving average (ARIMA) and k-nearest neighbors (kNN) in 1HA and 2HA solar power forecasting, especially after being optimized by the genetic algorithm (GA). Compared to traditional statistical models, AI models that only use endogenous data are less popular. The previous 80 min power values were used as the input to ANN models for 10MA–60MA forecasting to avoid intensive exogenous data while obtaining satisfactory accuracy [78].

A collection of research has been done to analyze the impact of combinations of endogenous and exogenous data on the forecast accuracy. Including more informative and well-selected input in the model will generate more accurate forecasts. Nonetheless, forecasting with only endogenous is still widely adopted for different purposes. Actually, most papers use traditional statistical models with endogenous input as benchmarks to validate the superiority of exogenous data-included models [79–82]. For example, [83] built a set of benchmark models, including ANN, support vector regression (SVR), and Gaussian Process (GP) models, solely using past GHI measurements. In some cases where the computational time is more important than the accuracy, exogenous data are not considered in the input. In [74], the power values of the last 10 min along with the corresponding 10 min from 24 h ago as well as a year ago were used to predict 1MA solar power for energy storage system control.

3.1.2. Meteorological measurements

The most natural way to extend the forecasting with endogenous data is to include meteorological measurements, since most weather stations and solar power plants are equipped with various types of meteorological measurement devices. Typical weather parameters include irradiance, temperature, wind speed, pressure, etc. For example, [84] used the maximum and minimum temperature to improve the ANN accuracy in GHI forecasting. It was found that more exogenous temperature from ancillary stations improved the forecast accuracy. [85] investigated the impacts of different weather parameter combinations

on the ANN accuracy and found that the input with all weather data (i.e., module temperature, ambient temperature, and irradiance) yielded the best forecast accuracy in 1HA–24HA PV power forecasting. In [86], a total of 74 inputs, including power, global normal irradiance (GNI), temperature, wind speed, humidity, pressure, and calendar features, were used in 15MA–60MA solar power forecasting, which found that the latest power and the time difference in respect of sunrise were the two most critical parameters in forecasting.

Forecasting with and without exogenous data was compared in the literature to show the effectiveness of involving exogenous data. For example, [79] compared ARMA with only power data and ANN with power and pressure, nebulosity, ambient temperature, wind speed, peak wind speed, wind direction, sunshine duration, relative humidity, and rain precipitations. The exogenous data improved the ANN model by up to 7.8% [82] and 9% [79] in 1HA and 24HA solar forecasting, respectively. Nevertheless, it was reported in some research that the benefit of weather parameters to the accuracy improvement in very short-term solar forecasting was limited. For instance, [87] found that weather parameters (i.e., temperature, humidity, wind speed, and cloud cover) were not able to improve the 5MA–60MA solar power forecasts. Therefore, it is suggested to carefully select weather feature combinations in specific forecasting tasks to avoid overconfidence in the merits of meteorological features.

3.1.3. Numerical weather prediction

AI techniques are used as post-processing approaches to enhance NWP forecasts. [88] found that variables related to lagged observations were more important for shorter forecasting horizons while the importance of NWP forecasts increases in longer horizons. NWP forecasts from ALADIN-France were combined with solar GHI in ANN models to provide 1HA GHI forecasts with a 0.7% nRMSE improvement [89]. The authors expected that the results could be further improved by better NWP forecasts.

It is observed that NWP improves longer-term solar forecasting more significantly. For example, [90] proved that the improvements brought by the European Centre for Medium-Range Weather Forecasts (ECMWF) were negligible in 3HA forecasting. However, the best 6HA forecasts were provided by the ANN models built with ECMWF outputs and the satellite GHI data. Moreover, a 42.9% improvement was obtained in 1DA GHI forecasting by feeding NWP forecasts from the North American Mesoscale Forecast System (NAM) into a long short-term memory (LSTM) model [91]. The importance of NWP forecasts from the Japan Meteorological Agency mesoscale model was assessed in [88] for different time horizons by their contributions to gradient boosted regression trees. It was found that the importance of some NWP forecasts, such as cloud cover indices and the western wind speed, increased with the time horizon. Nowadays, NWP models have a more rapid update rate and better granular spatio-temporal resolutions, which are encouraged to be involved in AI-based solar forecasting approaches [92].

3.1.4. Satellite data

The advantage of utilizing satellite images in solar power forecasting is their capability of monitoring the amount and movement of clouds [93]. Satellite images provide useful information for solar forecasting, such as cloud movements, coverage, and derived irradiance. For example, [94] computed cloud velocity and fraction from Geostationary Operational Environmental Satellite (GOES) images, which was input into ANN models for 30HA–120HA GHI forecasting. The improvements over persistence ranged from 5% to 25% according to different time horizons. In [95], the cloud cover index (CCI) was derived and predicted by the self-organizing map and the exponential smoothing state space model from the Japan Meteorological Agency Himawari-7 satellite images. Then, the CCI forecasts and historical GHI were fed into an ANN model for 1HA GHI forecasting with superior accuracy.

Satellite data is always combined with NWP forecasts in AI models for better accuracy. [96] predicted irradiance from satellite- (i.e., GOES-9) and NWP-derived (i.e., the Conformal-Cubic Atmospheric Model) irradiance values by the generalized additive model (GAM), which achieved a 3.6% lower RMSE, compared to the GAM model with only satellite-derived GHI. The most correlated features were selected from 30 satellite-derived irradiance and combined with ECMWF forecasts for 6HA GHI forecasting with the best accuracy [90]. A deep neural network (DNN) was used to forecast 1HA–6HA GHI at 25 locations in the Netherlands from the METEOSAT satellite image-derived irradiance and ECMWF forecasts. The developed method showed equal or better accuracy than the benchmarks trained with ground measurements, which provided a successful alternative to local telemetry-based forecasting.

However, the accuracy of satellite image-based solar forecasting is highly affected by the spatial resolution and time horizon. [93] found that the irradiance forecasting based on cloud impact factors derived from satellite images would have larger errors for smaller target areas or longer time horizons. Specifically, the SVR forecasting coefficient of determination (R^2) decreased from 0.91 to 0.79 when the radii of areas reduced from 35 km to 0.9 km, and the RMSE of 300MA forecasting could be 5 times larger than that of 15MA forecasting. Therefore, forecasters should be aware of the risk of accuracy deterioration in shorter-term forecasting for solar sites with fine spatial-granularity.

3.1.5. Sky images

Unlike NWP and satellite images, sky images are especially helpful in intra-hour solar forecasting. Compared to other ground measuring devices, sky imagers are more cost-efficient. Two research groups at the University of California, San Diego have done abundant work on sky image-based solar forecasting. In their early work, cloud coverage indices were calculated from sky images by the red blue ratio (RBR) thresholding and fed into ANN models together with measured GHI to forecast GHI [97]. Then, cloud fractions of grid elements along the cloud incoming direction were extracted to forecast intra-15MA direct normal irradiance (DNI) with encouraging accuracy [98]. The same cloud fraction features were combined with lagged DNI values to forecast 5MA and 10MA DNI [99], 5MA, 10MA, 15MA PV power [100], and 10MA irradiance ramps [101] with ANN models. [102] found that the mean, standard deviation, and entropy of red, blue, green, and RBR values from sky images provided limited improvements to kNN models, compared to the lagged irradiance features, especially for GHI forecasting. As they explained, one possible reason was that the image-derived parameters are only relevant when clouds are present. However, the same pixel statistics were found to benefit the 5MA–20MA DNI interval forecasting [101]. This might be due to that more advanced ML models, such as SVR and ANN, were used for the forecasting tasks.

Similar features were also utilized to help very short-term solar forecasting by other researchers. For example, sky image pixel RBR statistics, i.e., mean, standard deviation, and entropy, were extracted and combined with meteorological measurements for 1HA GHI forecasting in different frameworks [103–105]. [106] trained ANN models with sky image-derived cloud coverage, average, mode, median, and standard deviation of the red-green–blue (RGB) channels, and other parameters. These features were found to be especially helpful for intra-30MA forecasting. A total of 26 features were extracted from each sky image and selected by the principal component analysis in [107]. Then, these features together with solar azimuth, solar elevation, and the clear-sky irradiance were used to train ANN models, which showed better GHI forecasting performance than satellite image-derived models.

Nevertheless, the human-defined features may lose important information, which cannot well represent sky images or not optimal for AI models. With the advent of deep learning, sky image features can be extracted automatically by encoding techniques. For example, [108] constructed sky image features by building a convolutional neural network (CNN) that was used for cloudiness condition classification. Similarly, [109] first extracted sky image features from a 3D classification CNN model, which were used in an ANN model for 10MA GHI

forecasting. The automatically extracted features were found to improve the forecast accuracy comparing to human-crafted features.

Recently, researchers have investigated the possibility to predict power and irradiance directly from images. For example, [110] developed an end-to-end framework, called the SolarNet, which took one sky image as input and forecast 10MA GHI without any feature engineering. [33] proposed a similar approach, but included a sequence of sky images for contemporaneous PV power estimation. Then, the input and output of this framework were optimized by including lagged PV measurements, which outperformed the persistence by around 16% [111,112]. It is expected that by applying advanced deep learning architectures, the accuracy of sky image-based solar forecasting could be further improved.

3.1.6. Spatio-correlated hierarchical information

Solar irradiance or power is highly spatially and temporally correlated. Therefore, information within the temporal and spatial hierarchies is useful to improve AI model performance in solar forecasting. For example, solar irradiance or power measurements from neighbouring PV systems are input into AI models to capture the spatial patterns. [113] compared different combinations of 4 nearby PV systems and their meteorological measurements as exogenous inputs to ANN models for solar power forecasting. It was found that considering more information on neighboring distributed PV systems and meteorological measurements could enhance forecast accuracy. [114] considered a significantly larger number of PV systems in their kNN forecasting model. A "PV-Sensor Field", composed of 202 distributed rooftop PV systems, was used to capture the cross-correlations among PV-systems that were influenced by the same cloud sequentially. The developed method indicated superior performance than the persistence during highly variable days in 5MA–8MA forecasting. [84] improved ANN model performance by adding 30 meteorological stations' weather measurements in solar irradiation forecasting. [115] compared four data-driven models, i.e., ANN, SVR, boosted regression tree (BRT), and least absolute shrinkage and selection operator (LASSO), which forecasted 30MA–120MA GHI with the nearby 65 sites. BRT was reported to have the best performance with the help of neighboring solar irradiance data. Improvements associated with the spatio-temporal data is affected by the time horizon. It was found in [116] that the ANN model trained with data from the neighboring 65 stations reduced errors in 1HA–3HA forecasting, while the model that used only initial data from the station for which the prediction was made generated the best 4HA–6HA forecasts.

The derived neighboring solar irradiance or power is also beneficial to the onsite forecast accuracy. For example, [117] investigated the effects of inclusion of the satellite-derived GHI with spatio-temporal correlations in ANN and SVR models for 1HA–3HA GHI forecasting, and achieved significant improvements. Physically derived PV power was derived for 3 nearby solar farms and was used in conjunction with ECMWF predictions to train gradient boosting machine (GBM) models for PV power forecasting [118]. In a hierarchical system, the lower-level information provides details for higher-level forecasting. For example, solar power at each inverter was first predicted by ANN and SVR models and then summed up to obtain the forecasts for the entire 500 kW solar plant [119]. Results showed that the forecasting mean percentage error was reduced by this simple hierarchical operation.

3.2. Artificial intelligence-based forecasting algorithm

Most AI-based solar forecasting models are ML models. ML models learn relations between inputs and outputs from data even if the representation is impossible. Compared to NWP models, ML models have significantly more frequent update rate and are more accurate for very short-term solar forecasting. The most widely used ML models for solar forecasting are kNN, ANN, SVR, random forest (RF), GBM, and deep learning models. Forecasting accuracy is related not only to the learning capability of ML models, but also other factors, such as solar time series forecastability, meteorological conditions, pre- and post-processing, etc. Comparisons of ML models should be made with the same condition. This is commonly ensured in case studies within each paper, but relatively challenging to fulfill across multiple papers. Therefore, the evaluation metrics listed in the tables in this section are for reference purposes. This will be further detailed in the discussion section.

3.2.1. K-nearest neighbors-based solar forecasting

kNN is a non-parametric method, which is one of the simplest ML algorithms. In kNN, solar forecasts are calculated by the weighted average of the k-nearest neighbors that are determined by the distance. kNN could be used as the forecasting engine to replace the persistence model. For example, [102] predicted 5MA–30MA GHI, DNI, and their intervals from local telemetry and sky images with kNN, which showed a 10%–25% improvement compared to the persistence method. kNN models also indicated better accuracy than NWP models, such as reported in [123]. Although kNN might perform well [120,122,121,74], it was not as competitive as more advanced ML models in most cases, as shown in Table 4. For example, both GA optimized ANN and ANN beat kNN for 1HA–2HA PV power forecasting in [77,100]. Therefore, kNN was mainly used as a benchmark or for comparative analysis [124,126,127,74]. [127] included kNN in a pool of 11 ML models for 12HA PV power forecasting and found that kNN was worse than SVR and decision tree (DT) models. SVR, RF, and kNN were compared in [126] for 1HA GHI forecasting, where kNN was less accurate. However, combined with other models, kNN performance could be enhanced. For example, [129] ensembled GB, ANN, kNN, and SVR by least squares, which improved the 1DA PV power forecast accuracy by 5%. As an instance-based learning method, kNN is easy and fast to implement and works well on small datasets.

3.2.2. Artificial neural network-based solar forecasting

ANN is the most popular solar forecasting model algorithm. ANN captures relations between inputs and output(s) with multiple interconnected layers of neurons. The weighted input neurons are connected to the output neurons through activation functions. The training processing is to update the weights and biases so that the objective function, i.e., errors in solar forecasting, is minimized. Based on the network architecture, ANN models can be further divided into feedforward neural

Table 4
Representative kNN solar forecasting papers.

Reference	Location	Horizon	Comparison
[77]	US	1HA–2HA	GA-ANN>ANN>kNN>ARIMA>persistence
[120]	China	1DA–3DA	kNN and ANN > SVR, MLR, and persistence
[100]	US	5MA–15MA	ANN>ARIMA>kNN>persistence
[102]	US	5MA–30MA	kNN>persistence
[121]	US	15MA–2HA	kNN, ANN>persistence
[122]	US, Denmark, Italy	1DA	kNN>ANN, SVR>persistence
[123]	US	NA	kNN > physical model
[124]	China	1DA	MARS>kNN>SVR>ANN>CART
[125]	NA	1DA	GBM>RF>kNN>DT
[126]	Spain	1HA	SVR>RF>kNN
[127]	NA	12HA	Conflicting results
[128]	US	5MA–20MA	kNN > persistence
[129]	Italy	1DA	Conflicting results
[130]	Taiwan	1HA	kNN-ANN > kNN
[131]	Saudi Arabia	1HA	ANN>kNN>DT>SVR
[132]	US	5MA–30MA	GBM>kNN>persistence
[74]	US	1MA	kNN>SVR>RF

network (FFNN), radial basis function neural network (RBFNN), multilayer perceptrons neural network (MLP), and extreme learning machine (ELM). Other types of ANN, such as the recurrent neural network (RNN, mainly the LSTM models) and CNN will be discussed in deep learning section. These four groups of ANN models are not mutually exclusive and their representative papers are listed in Table 5.

Specifically, an FFNN is an ANN architecture, wherein connections between neurons do not form a cycle, which is different from RNNs. FFNNs are sometimes called BPNN or FFNN with back propagation. Actually, back propagation is a training algorithm, which takes the feedforward values to calculate errors that are propagated to previous layers through the chain rule. FFNNs are widely used in solar forecasting. For example, [135] investigated the impact of input neurons on the FFNN performance in solar radiation forecasting and found that using 6 neurons in the input layer could receive more accurate forecasts. [136] developed a three-layer FFNN with 28, 12, and 11 neurons in the input, hidden, and output layer, respectively. The proposed architecture was determined by the selected input, designed output, and an empirical formula, which provided accurate 24HA PV power output forecasts. A similar approach was used to determine the input and output neurons in [137], while a trial-and-error method was used to determine the hidden layer neurons. [77] used an automatic way to optimize the FFNN architecture for 1HA–2HA PV power output forecasting. Specifically, GA was used to optimize the numbers of input and hidden layer neurons, which improved the performance by 30%.

Table 5
Representative ANN solar forecasting papers.

Reference	Location	Type	Horizon	Evaluation metrics
[77]	US	FFNN	1HA–2HA	RMSE= (116.54–162.37) kW
[120]	China	FFNN	1DA–3DA	MAPE= (17.31–49.87)%
[133]	Algeria	FFNN	NA	RMSE = 64.34 W/m^2
[134]	Singapore	FFNN	NA	MAPE = 6.03%
[121]	US	FFNN	15MA–2HA	RMSE=(41.9–103.9) W/m^2
[129]	Italy	FFNN	1DA	nMAE=(1.29–5.16)%
[94]	US	FFNN	30MA–4HA	RMSE=(49.79–81.26) W/m^2
[124]	China	FFNN	1DA	RMSE=(120.2–135.6) W
[135]	Turkey	FFNN	NA	nRMSE = 3.58%
[127]	NA	FFNN	12HA	MAPE=(0.59–2.33)%
[136]	US	FFNN	24HA	MAPE=(6.78–7.67)%
[137]	US	FFNN	24HA	MAPE= (10.06–18.89)%
[95]	Singapore	FFNN	NA	nRMSE= (28.19–43.69)%
[138]	Portugal	FFNN	1–3DA	FF = 28.2%
[139]	Germany	FFNN	1DA	23%
[131]	Saudi Arabia	MLP	1HA	RMSE=(43.75–47.44) W/m^2
[140]	France	MLP	1DA	RMSE = 3.73 MJ/m^2
[141]	NA	MLP	1DA	nRMSE = 7.78%
[142]	NA	MLP	1DA–3DA	RMSE=(42.29–84.65) W/m^2
[143]	France	MLP	24HA	RMSE=(33.10–92.55) W/m^2
[144]	Italy	MLP	24HA	R^2=(95–98)%
[145]	France	MLP	6HA	NA
[146]	The Netherlands	MLP	1DA	rRMSE = 31.31%
[147]	Saudi Arabia	RBFNN	NA	R^2=98.80%
[148]	China	RBFNN	24HA	MAPE = 9.45%
[122]	US, Denmark, Italy	RBFNN	1DA	nRMSE= (5.99–11.41)%
[149]	India	RBFNN	NA	RMSE = 2.30 kWh
[150]	Malaysia	ELM	1HA and 1DA	nRMSE= (13.83–21.84)%
[151]	India	ELM	15MA–60MA	MAPE=(1.44–4.19)%
[81]	Multiple countries	ELM	15MA–24HA	RMSE=(74.5–197.3) W/m^2
[152]	India	ELM	15MA–1DA	MAPE = 1.244%
[153]	US	ELM	NA	MAPE = 2.55%
[154]	NA	ELM	5MA	RMSE = 0.188 W/m^2
[155]	Jordan	ELM	25HA	RMSE = 15.07 kW
[156]	Spain	ELM	NA	RMSE= (67.28–107.19) W/m^2
[157]	Singapore	ELM	5MA	NA
[158]	China	ELM	NA	RMSE=(3.83–5.86)%

RBFNN is a type of ANN that has three layers with a nonlinear radial basis function (RBF) as the activation function. RBFNNs are popular in solar forecasting due to their better approximation capabilities, simpler network structures and faster learning algorithms. For example, [147] compared different RBFNNs in the daily solar radiation forecasting using meteorological data and found that RBFNN was better than MLP. In [148], RBFNNs with multiple output neurons were developed for 24HA PV power output forecasting under various weather conditions. Different numbers of neurons in the hidden layers were tested to ensure accurate forecasts. [149] developed RBFNNs to forecast daily power generation of PV plants in 26 different Indian cities, which provided more accurate forecasts than polynomial regression.

MLP is a special class of FFNN, which usually contains multiple hidden layers. With more hidden layers, MLP is able to learn more complex patterns in the data for solar forecasting, which also makes the model optimization more challenging. Therefore, optimizing the topology and hyperparameters is critical when applying MLP in solar forecasting. For example, [131] proposed an MLP with 2 hidden layers with 7 and 5 neurons, respectively. The numbers of neurons at four different layers were selected by adding more neurons and keeping track of the RMSE performance, until the error is minimized over the training data. The MLP model was found to outperform SVR and kNN in case studies. Cross-validation was used in [144] to obtain the best MLP configuration for 24HA solar irradiance forecasting. The best MLP had two hidden layers, with 11 neurons in the first hidden layer and 17 neurons in the second layer. [140] did a more comprehensive exploration of the MLP hyperparameter settings, which included the number of neurons in the input layer, hidden layer, activation functions, and optimization functions. The developed MLP provided more accurate forecasts than ARIMA, Bayesian inference, Markov chains, and kNN. A more advanced optimization algorithm, GA, was used in [101] to optimize the MLP input, number of layers, and number of neurons in each layer, which achieved 6.0%–11.3% improvements in 10MA GHI and DNI forecasting. [145] used MLP as the forecasting model to study the forecasting uncertainties associated with the measurement error, the time series variability, the machine learning model, and the error related to the horizon.

ELM is a special type of single layer FFNN. Compared to ANNs that are trained with back propagation, ELM randomly selects the hidden layer weights and biases and determines the output weights by the Moore–Penrose generalized inverse, so that the training speed is thousand times faster. Due to the outstanding performance in terms of learning speed and accuracy, ELM is widely used in solar forecasting. For example, the training time of ELM was 0.2028–0.3432 s, which was less than that of ANN (0.7225–0.7405 s) and SVR (2.9140–3.8612 s) in [150]. [156] explored using several ML models, including SVR, MLP, GP, and ELM, to forecast solar radiation from satellite data, where the ELM was the best. The ELM forecasting errors were also smaller than ANN and SVR in the case studies of this research. The ELM performance was boosted by 3 optimization algorithm, i.e., particle swarm optimization (PSO), Craziness PSO, and accelerated PSO (APSO), which selected appropriate values of weights. It was found that APSO-ELM achieved better performance than standard ELM and ELMs with other optimization methods [151]. [81] optimized the ELM model by selecting the best input features using mutual information criteria, which had the advantage of achieving good performance in terms of accuracy within an extremely fast computational time. Other ELM papers that used similar techniques are listed in Table 5 without discussion.

3.2.3. Support vector regression-based solar forecasting

SVR is a kernel-based ML algorithm, which constructs a hyper-tube in a high-dimensional space that best approximates the continuous-valued function, while balancing model complexity and prediction error [169]. SVR is another popular algorithm used in solar forecasting. Based on the kernel function, SVR can be divided into SVR with linear, polynomial, or RBF kernel, among which RBF is the most widely-used kernel. Table 6 summarizes the latest representative papers that used SVR in solar forecasting. A more comprehensive review of SVR and SVM applications can be found in [38].

Linear kernel-based SVR was found to provide more accurate forecasts than other ML models in some case studies. [126] adopted linear kernel function in SVR for 1HA GHI forecasting. Compared to benchmarks, such as RF and kNN, SVR achieved the best forecast accuracy with MAE ranging from 49 W/m^2 to 64 W/m^2.

RBF is generally more flexible than linear and polynomial functions, therefore, is more popular. For example, [160] developed SVR with an RBF kernel to forecast 15HA–33HA power output from NWP outputs for a 1 MW PV power plant in Japan. Case studies with 2 years of data showed that the developed SVR model reduced forecasting errors by more than 10%, compared to the persistence method. It was also found that the SVR with cloudiness was improved by almost 33% and 80% based on RMSE and MAPE, respectively. [161] also used RBF as the kernel function in SVR for solar forecasting. The SVR was solved by using Lagrange multipliers, which achieved better accuracy than AR and ANN in 1HA–3HA forecasting for 9 months data in 2005. Various preprocessing techniques were applied to improve the accuracy. A 2D data representation and transmissivity-based normalization was suggested by the results. Similarly, SVR with RBF was compared with hidden Markov model (HMM) in 5MA–30MA solar irradiance forecasting, which showed better forecasts with accuracy larger than 90%. SVR was used as a blending model to post-process forecasts from satellite-based model, NWP model, smart persistence, and a hybrid satellite-NWP model in [159]. It was found that the blending model improved GHI forecasting by 17%. [168] found the annual value of SVR forecasting increased 22.32€for each 1 kWh improvement in RMSE.

Weather-aware forecasting that has one SVR for each weather condition is able to further improve the forecast accuracy. [162] applied bi-model SVRs with RBF that are separated based on the cloudiness for 5MA–15MA solar irradiance forecasting. The developed method showed superior performance in terms of forecast accuracy and ramp-down event detection. [165] used 3 SVR models with RBF to forecast 1HA PV power in different weather conditions that were clustered by the k-means method. It was found that SVRs generated better forecasts, compared to ANN that used the same methodology. [163] divided the scenarios using self-organizing maps and built an RBF-based SVR for each scenario. The parameters of SVRs were optimized by PSO. The optimized SVRs were more accurate than the random walk method, exponential smoothing method, and ARIMA method.

As a black-box model, it is always challenging to interpret results from SVRs. In [166], structural variable selection methods were introduced to SVR to consider the heredity principle and sparsity in the data. The SVR consistency was guaranteed by selecting the RBF kernel parameters using an information criterion. The developed SVR model improved the forecast accuracy by 5.06%–70.03%.

Table 6
Representative SVR solar forecasting papers.

Reference	Kernel	Location	Horizon	Evaluation metrics
[126]	Linear	Spain	1HA	MAE=(49–64) W/m^2
[159]	Linear/ RBF	Multiple sites	6HA	SS = 17%
[160]	RBF	Japan	15HA–33HA	MAPE = 29.53%
[161]	RBF	US	1HA–3HA	MAE=(34.37–92.57) W/m^2
[162]	RBF	Taiwan	5MA–15MA	MAPE=(20–22)%
[163]	RBF	US	1HA	nRMSE=(22.5–45.0)%
[164]	RBF	Australia	5MA–30MA	NA
[165]	RBF	Korea	1HA	RMSE=(49.26–62.57) W/m^2
[127]	RBF	Italy	12HA	MAPE=(0.83–2.59)%
[166]	RBF	China	NA	RMSE=(14.20–17.85) W/m^2
[167]	RBF	China	1DA	nRMSE = 2.96%
[168]	RBF	Iberia	1DA	nRMSE = 22.54%

3.2.4. Ensemble learning-based solar forecasting

Ensemble learning is a methodology that combines multiple learners to achieve better performance. There are different ways to categorize ensemble learning methods. For example, ensemble learning can be divided into homogeneous ensemble methods and heterogeneous ensemble methods. Base learners of the former group are identical, such as classification and regression trees (CARTs), while the latter group of methods have distinct base learners. According to the strategy, ensemble learning can be grouped into parallel ensemble and sequential ensemble. Parallel ensemble requests base learners to be independent, such as RF, while sequential ensemble relies on the dependence of base learners, such as GBM. In this section, the most popular ensemble methods are introduced, including RF, GBM, and hybrid models. Different types of ensemble methods are summarized in Table 7.

RF is a kind of regression tree model that has a multitude of decision trees, each of which is built with a subset of training data with replacement. The final forecasting is obtained by taking the conditional mean of forecasts from decision trees. Therefore, forecasting results are more robust and accurate. For example, RF was found to outperform ANN in forecasting three components of solar irradiance in a location in France [170]. The nRMSE of RF ranged from 19.65% to 27.78% for 1HA to 6HA GHI forecasting. [171] compared 11 ML models in 1HA–6HA solar irradiance forecasting with different variability levels. RF provided satisfactory forecasts, especially for locations with higher variabilities. A total of 68 ML models were compared in [173] in 1HA GHI forecasting, where the RF models was found to be competitive in most cases. [184] included RF in the 1HA solar forecasting benchmarks using several publicly available datasets.

Table 7
Representative ensemble learning-based solar forecasting papers.

Reference	Method	Location	Horizon	Evaluation metrics
[170]	RF	France	1HA–6HA	nRMSE= (19.56–27.7)%
[125]	RF	NA	1DA	RMSE = 0.083–0.098
[171]	RF	Multiple sites	1HA–6HA	nRSME= (18.97–48.34)%
[172]	RF	US	3HA	MAE = 2277.2 kJ
[173]	RF	US	1HA	nRMSE= (25.83–33.44)%
[88]	GBM	Japan	1HA–6HA	RMSE= (9.48–13.17)%
[174]	GBM	US	1DA	NA
[125]	GBM	NA	1DA	RMSE = 0.082–0.101
[118]	GBM	NA	1DA	NA
[175]	GBM	Portugal	6HA	SS=(1.4–5.9)%
[176]	GBM	Singapore	1DA	SS = 4%
[177]	GBM	Portugal	24HA	SS=(12.06–13.11)%
[178]	GBM	US	1DA	nRMSE= (6.96–7.72)%
[132]	GBM	US	5MA–30MA	SS=(13–24)%
[172]	XBM	US	3HA	MAE = 2190.9 kJ
[179]	ANN	Australia	1DA	MAPE = 3.1%
[180]	SVR-RF	Australia	NA	RMSE = 0.0725
[181]	RLS-GBM	Portugal	1HA–6HA	SS=(8–12)%
[103]	ANN-SVR-RF-GBM	US	1HA	nRMSE = 9.74%
[182]	SVR-RF	Multiple sites	1HA	nRMSE= (8.18–34.35)%
[183]	SVR-ANN	Australia	1DA	RMSE = 0.2133 kW

GBM is a sequential ensemble method that also mainly uses decision trees as base learners. GBM works in the form of a numerical optimization problem where the objective is to minimize the loss of the model by adding weak learners using a gradient descent-like procedure. [132] found that GBM outperformed kNN and the persistence in 5MA–30MA GHI and DNI forecasting. [88] used GBM to forecast 1HA–6HA power output for 42 rooftop PV systems in Japan. In addition, feature contributions were extracted from GBM models to analyze the feature importance. GBM was also reported to perform well for longer-term forecasting. In [178], GBM beat SVR in 1DA solar forecasting. GBM was also suitable and widely used for probabilistic forecasting. For instance, a GBM model was fit to predict each quantile ranging between 5% and 95% with 5% increments in 6HA PV power output forecasting, which achieved 1.4%–5.9% improvements [175]. Similaly, GBM was used in [176] for solar irradiance quantile forecasting, which improved probabilistic forecasts by 4%. In addition, GBM was used to convert gridded solar irradiance that was predicted by NWP to probabilistic PV power output, which received more improvements (more than 12%) than directly predicting quantiles [177]. In the Global Energy Forecasting Competition (GEFCom), GBM was the most popular forecasting engine [185,118,186]. It is surprising that two other boosting algorithms, i.e., adaptive boosting and extreme gradient boosting (XGB), were rarely applied in solar forecasting.

Comparisons were also made among ensemble learning methods. For example, [125] found that RF performed better than GBM with limited data, but GBM was more accurate with more data in 1DA PV power output forecasting. [174] tested various ensemble learning methods, including GBM, RF, and other tree methods, in solar forecasting of Oklahoma gridded sites. Results showed that GBM had the lowest MAE. Three ensemble learning models, i.e., RF, GBM, and XGB, were compared in [172]. The more advanced ensemble learning method, XGB, outperformed others in solar forecasting tasks.

Stacking is an ensemble learning technique that combines multiple models, which are trained based on a complete training set. Another model is trained on the outputs of the base level model as features. These models are sometimes referred as hybrid models. Stacking methods take advantage of multiple ML algorithms, which are more robust and accurate than any single ML models. For example, [180] ensembled SVR results using RF for PV power output forecasting, which reduced forecasting error by 5%–9%. [185] stacked RF and GBM models to predict the probability distribution of PV power, which achieved a high ranking in the GEFCom. Four ANNs were used as base learners in [187] to forecast power output for a 10 Wp PV system. [182] averaged several models, including SVR and RF, in 1HA GHI forecasting, and found the ensemble models performed slightly better than single algorithm models. A set of ANN, SVR, RF, and GBM models were used as base learners, which were ensembled by another ML algorithm for 1HA GHI forecasting [103,105,104]. The ensemble model outperformed any base learner. Since base models are usually heterogeneous, therefore this type of ensemble has various combinations that are hard to list exhaustively in this review. More ensemble solar forecasting papers can be found in [188].

3.2.5. Deep learning-based solar forecasting

Deep learning, a special branch of ML, is an emerging solar forecasting technique in recent years. Different from shallow learning methods, deep learning is a promising alternative to further improve forecast accuracy by its unsupervised feature learning, strong generalization capability and, and big-data training [40]. The most prevalent deep learning architecture in solar forecasting are LSTM and CNN.

LSTM is a category of deep learning architectures that belongs to RNN. Different from FFNN, LSTM has feedback connections, allowing information persists, which brings breakthroughs and makes LSTM a natural choice for time series analysis, such as solar forecasting. LSTM was reported to outperform FFNN and SVR in forecasting the power output of a Canadian PV plant [190]. Similarly, LSTM was found to be 18.34% more accurate than FFNN in [91]. The robustness of LSTM was verified in [194] by comparing LSTM to FFNN and GBM in 1DA GHI forecasting. [189] compared three deep learning models, i.e., LSTM, deep belief networks (DBN), and autoencoded-LSTM (Auto-LSTM), with MLP and a physical model. It was found that Auto-LSTM was the best-performing model in up to 2DA PV power forecasting. [191] found that LSTM performed better than another popular deep learning architecture, CNN, by 9%. Forecast accuracy could be significantly improved by applying LSTM to some datasets. For example, [192] reported that 50.90%–68.89% forecasting skill scores of the LSTM model were achieved over the smart persistence.

CNN is another popular deep learning structure for visual imagery analysis. CNNs capture spatial and temporal dependencies in images through filters. In solar forecasting, CNNs are used to provide forecasts through processing time series, images, or video streams. For example, [195] improved 15MA–2HA PV power forecast accuracy by up to 61.42% with a 1D CNN that took time series data as inputs. [110] developed a deep learning network for solar forecasting, called the SolarNet, which took only one sky image as input and improved 10MA GHI forecast accuracy by 11%. In the following paper, they extended the SolarNet to intra-hour forecasting, which improved the accuracy by 25% [196]. [111] utilized both sky image video and time series data in their deep learning network, where input and output are optimally determined. The combination of CNN and LSTM is also a promising way to further improve solar forecast accuracy. For example, [198] confirmed that by stacking CNN and LSTM, the performance of the model could be enhanced. [197] also stacked CNN and LSTM sequentially, which improved 1DA forecast accuracy compared to CNN and LSTM. A similar CNN-LSTM architecture that took multi-variant input was proposed in [199], which outperformed CNN and LSTM models. A list of deep learning forecasting models are collected and grouped in Table 8.

3.3. Summary

In this section, AI-based solar forecasting is taxonomically reviewed by input data and ML algorithms. The efficient techniques that ensure a successful AI forecasting model lie in one or more of the following steps:

- Feature engineering. The incorporation of data from diverse sources, such as endogenous time series, meteorological data, sky images, shadow images, and satellite observations, is important. Taking full advantage of these data is even more critical. For shallow ML, proper

Table 8
List of representative deep learning-based solar forecasting papers.

Reference	Method	Location	Horizon	Evaluation metrics
[189]	LSTM	German	24HA–48HA	RMSE = 0.0713
[190]	LSTM	Canada	NA	RMSE = 0.086
[191]	LSTM	Japan	1MA–10MA	SS = 21%
[91]	LSTM	Cape Verde	1DA	RMSE = 76.245 W/m^2
[192]	LSTM	Multiple sites	1DA	SS=(50.90–68.89)%
[193]	LSTM	Egypt	1HA	RMSE=(82.15–136.87) W/m^2
[194]	LSTM	Multiple sites	1DA	SS = 52.2%
[195]	CNN	Belgium	15MA-2HA	SS=(49.10–61.42)%
[111]	CNN	US	15MA	SS=(15.7–16.3)%
[112]	CNN	US	NA	SS = 17.1%
[110]	CNN	US	10MA	SS = 11.88%
[196]	CNN	US	10–60MA	SS = 25.14%
[197]	CNN-LSTM	NA	1DA	MAE=(16.53–130.94) W/m^2
[198]	CNN-LSTM	Korea	1HA–6HA	MAPE=(13.42–37.83)%
[199]	CNN-LSTM	Australia	1DA	MAPE=(2.2–11.2)%
[200]	CNN-LSTM	US	1HA–4HA	MAPE=(14.6–21.4)%

feature engineering, including feature extraction, selection, and construction, should be performed automatically with advanced algorithms.

- Model optimization. The model selection and optimization are required to obtain optimal forecasts, including the consideration of weather and calendar effects, hyperparameter selection, and architecture optimization. Due to the complexity and heterogeneity of forecasting tasks, these processes should also be conducted with advanced optimization techniques.
- Post-processing. Outputs of AI-based forecasting models should be processed to meet the operational requirements, including spatial or/and temporal reconciliation, clipping, and irradiance to power conversion. These processes should follow the requirements of specific power system operators, or else the application value of the forecasts will be suspicious.

Future trends of the AI-based forecasting are identified by the latest papers and summarized as:

- Sky, shadow, and satellite images. Different from time series, solar forecasting with image data is still less researched. Inclusion and optimization input combinations from various sources are expected to further improve AI-based forecasting performance.
- Deep learning. Compared to shallow ML, the application of deep learning in solar forecasting is still at its early stage. The development of deep learning techniques is expected to capture spatial and temporal patterns in multiple data sources in solar forecasting data in a more efficient way. Deep learning-based solar forecasting could advance or at least be competitive with shallow ML methods.
- Probabilistic forecasting. Compared to load and wind forecasting, solar forecasting in probabilistic form is also lagging behind. Probabilistic solar forecasting will help power system operators better manage uncertainties associated with forecasts, therefore, becoming an emerging topic and will be continuously under investigation.

4. PV Array and Fault Detection

Detection in solar energy is to identify unknown information. Detection tasks in solar energy mainly include PV array detection and PV fault detection, which are the focus of this section. Different from forecasting that generates continuous outputs, detection belongs to classification problems that output discrete results. Therefore, most papers use classification evaluation metrics, including (i) threshold metrics, such as accuracy, precision, recall, F1 score, and the Jaccard index, (ii) ranking metrics, such as the receiver operating characteristic (ROC) curve, and (iii) probabilistic metrics, such as the log loss and the Brier score. The definitions and formulations of detection metrics can be found in [201–203].

4.1. Array detection

With the growing penetration of distributed PV, there is a strong interest among government and utility decision-makers in obtaining detailed information about rooftop PV [204]. This is due to that the key data of distributed PV systems maybe only partially known or completely absent [214,215], which is different from large utility-scale PV plants whose specifications are known by the commercial forecaster and system operator. Even though utilities typically require permits and enforce interconnection requirements for installing these behind-the-meter (BTM) solar generation [216] to reduce risks, the number of unregistered or unknown BTM PV systems is significant for a variety of reasons. For example, the installations may be made prior to regulations being enacted. Not all the BTM PV owners are aware of the rules, or some owners want to avoid permitting fees. There is also the possibility of having discrepancies between the reported and the installed configurations [217]. In recent years, AI methods make it possible to non-intrusively detect PV installations through remote sensing data. Table 9 lists PV array detection papers that rely on AI methods.

Like AI research in forecasting, shallow ML was first adopted, which needed to explore proper input features. For example, [204] first built an SVR model that learned from color, shape, and texture of 100 aerial images to detect PV objects. The classifier was able to identify if a PV object exists in an image by a 94% accuracy. Different features were extracted from aerial images, including raw pixels, local color statistics, and textons to detect the PV pixels [205]. It was found that the combination of local color and texture yielded the best detection based on the ROC curve. Then, the pixel means and variances of different sizes of windows were used as input to RF models for PV pixel detection in [206]. The detection accuracy was 90%, which was satisfactory considering the large dataset (i.e., 135 km^2 images).

With the prevalence of deep learning, PV array detection accuracy is significantly improved. Deep learning models, mainly CNNs, learn patterns and features from image automatically without heavy feature engineering. A CNN with only 3 convolutional layers achieved 98% accuracy by learning from 3,347 low-quality Google satellite images [208]. [209] developed a faster region-based convolutional neural networks (RCNN) based on only 800 200×200 images, which successfully detected PV objects with a 0.9299 precision. A 12-layer CNN model was reported to perform better than an RF model with fewer false alarms. The precision of CNN model is 70%, which was a significant improvement compared to the 10% precision of an RF model [207]. A few advanced deep learning architectures were applied in PV array detection. [210] developed the Visual Geometry Group (VGG) model for PV detection and found that transferring the pre-trained weights to a CNN, i.e., VGG, did not help improve the detection accuracy. The VGG model with random weight initialization achieved the best precision. [212] advanced PV array detection from PV object classification to PV pixel semantic segmentation. A SegNet model was developed based on 2 million images and outperformed the VGG model.

Before applying PV array detection models to large-scale real-world detection, a model generalization should be first evaluated, especially with the fact that there is wide variability in the visual characteristics of remote sensing imagery across different geographic locations. [218] trained CNNs with images of California, which was used to detect PV arrays in Connecticut. It was found that by using only a small number of images (i.e., 15 km^2) for fine-tuning, the detection precision was 0.88. [211] developed a CNN encoder-decoder architecture with three modules, which are feature refinement residual module, chained dilation attention module, and global channel attention module. The developed

Table 9
AI-based PV array detection papers.

Reference	Method	Input	Output	Evaluation
[204]	SVM	Aerial image	PV object	Accuracy = 0.94
[205]	RF	Aerial image features	PV pixel	ROC curve
[206]	RF	Aerial image features	PV pixel	Accuracy = 0.90
[207]	CNN	Aerial image	PV object	Accuracy = 0.90
[208]	CNN	Aerial image	PV object	Accuracy = 0.87
[209]	CNN (Faster-RCNN)	Aerial image	PV object	Precision = 0.93
[210]	CNN (VGG)	Aerial image	PV object	Precision = 0.95
[211]	CNN	Aerial image	PV pixel	Precision = 0.94
[212]	CNN (SegNet)	Aerial image	PV pixel	ROC curve
[35]	CNN	Aerial image	PV pixel	Precision = 0.93
[213]	CNN	Aerial image	PV object	Precision = 0.91

method learned from 223 images from 7 provinces of China and yielded better performance than several famous CNN architectures, including UNet and LinkNet. [219] applied CNNs for PV array detection on a large aerial imagery dataset comprised of two nearby U.S. cities. The results indicated that substantial quantities of local training data are needed to perform well in a new location, but far less than the size of a full training dataset. [35] developed a deep learning framework, called the Solar-Forecast, to automatically localize solar PV panels in the U.S. The CNN model was trained with 366,467 images sampled from over 50 cities/towns across the U.S., which achieved a precision of 93.1% in residential areas and a 93.7% precision in non-residential areas.

4.2. Fault detection

Faults in any components of a PV system can significantly affect the efficiency and security of the PV plant and the economic and reliable power system operations, which makes PV fault detection (FD) an important task in monitoring the PV system performance. Due to shading effects, module soiling, PV modules aging, and others, all the PV system components, such as modules, cabling, protections, converters, and inverters, are subject to failures [50]. The main catastrophic failures in PV arrays include: single-line to ground, line-to-line, line-to-line-to-ground, and 3-phase fault.[2] FD usually includes fault identification, diagnosis, and localization, which are used to identify, recognize, and localize faults in a PV system, respectively.

Different from FD methods that are based on PV electrical circuit simulation and electrical signal processing, AI-based FD learns from meteorological, electrical data, and images, which does not require prior knowledge of the PV system, such as system parameters and mathematical formulations. AI-based FD models are faster, more accurate, more flexible, and more scalable. Typically, there are two categories of AI-based FD methods: regression methods and classification methods. The former approaches identify faults by estimating system parameters using AI methods, which are used to compare with measurements. Classification methods directly classify fault types without knowing the monitoring parameters. Both approaches are effective and efficient in FD. Table 10 listed the recent two groups of AI methods for FD, where the output of fault type indicates FD classification methods and others are FD regression methods.

Theoretically, all the methods in Section 3 could be used in FD regression methods. However, most papers used ANN in this task. For example, a two-layer ANN was applied to predict the power from a PV system, from which the open-circuit voltage and short circuit current were derived and used to identify six fault types [222]. Similar work was shown in [223], where an ANN model was used to estimate the power output from irradiance, wind, and temperature. Then, system outages and other faults causing a reduction in power were detected from estimated and measured power. [220] optimized the ANN topology by GA, which was used to detect five fault types with a 90% accuracy.

Another similar way to use AI for FD is optimizing the mathematical model by numerical AI algorithms, which can better monitor the PV system. For example, an ANN model was used to correct the power estimates of a one diode model [226]. The improved power output estimation could be used to better identify specific faults.

To better identify and localize faults in PV systems, multiple AI models are required to monitor different signals. For example, twelve three-layer ANNs were used to monitor the 6 voltage and 6 current of a 3×2 PV system [227]. The method was able to identify the short-circuit location of PV modules in one string independently by the control rule. In [221], ANN models were found to have less learning time, high accuracy, and lower memory consumption while conducting FD based on a ZigBee wireless sensor network. The faulty operating state of a partially shaded PV module was detected by ANNs from irradiance and temperature. Four common faults of PV arrays, including the degradation fault, short-circuit fault, open-circuit fault, and partial shading condition, were diagnosed by an ELM model from key points of I-V curves [231]. A similar RBF ELM model was proposed in [232] to identify three types of faults.

Among FD classification methods, [229] identified and diagnosed 9 types of faults in a PV system using a 2–2-9 ANN architecture from irradiance and temperature with a 92.1% accuracy. [228] built a four-

Table 10
AI-based fault detection papers.

Ref.	Method	Input	Output	Evaluation
[220]	ANN	Irradiance, temperature	Power	Accuracy= (96.2–98.9)%
[221]	ANN	Irradiance, temperature	Power	Accuracy = 100%
[222]	ANN	Irradiance, temperature	Power	NA
[223]	ANN	Irradiance, wind, temperature	Power	NA
[224]	ANN	Weather forecasts, power	Power	nMAE= (2.17–15.78)%
[225]	ANN	Irradiation, temperature, current, voltage, power	Voltage, current	MAE=(0.0667–0.52) A
[226]	ANN	Irradiance, temperature	Power error correction	FP = (4.98–11.52)%
[227]	ANN	Irradiance, temperature	Voltage, current	NA
[228]	ANN	Irradiation, temperature, current, voltage, power	Fault type	Accuracy = 98.53%
[229]	ANN	Irradiance, temperature	Fault type	Accuracy = 92.1%
[230]	ANN	Irradiation, temperature, current, voltage	Fault type	Accuracy= (82.34–98.19)%
[231]	ELM	Current, voltage	Fault type	Accuracy= (99.77–99.40)%
[232]	ELM	Current, optimal fitness value, resistence	Fault type	Accuracy= (93.55–97.25)%
[233]	SVM	Power	Fault type	Accuracy= (68.0–75.8)%
[234]	SVM	Current, voltage	Fault type	Accuracy= (91.40–94.74)%
[235]	DT	Irradiation, temperature, current, voltage	Fault type	Accuracy= (93.56–99.99)%
[236]	DT	Irradiation, temperature ambient, power ratio	Fault type	Accuracy= (99.80–99.86)%
[237]	RF	Current, voltage	Fault type	Accuracy= (99.095–99.224)%
[238]	kNN	Irradiation, temperature, current, voltage	Fault type	Accuracy = 98.70%
[239]	LSTM	Current, voltage, power	Fault type	Accuracy= (97.66–100)%
[240]	CNN	Irradiation, temperature, current, voltage	Fault type	Accuracy= (88.852–98.774)%
[241]	CNN	Electrical time series graph	Fault type	Accuracy= (99.03–99.51)%
[242]	CNN-SVR	Aerial image	Fault type	Accuracy = 90.23%
[243]	CNN	Aerial image	Fault type	Accuracy>95%
[244]	CNN	Aerial image	Fault type	Accuracy = 98.7%
[245]	CNN	Infra-red image	Fault type	F1 = 0.65–0.69

Note: NA indicates that the overall performance of the detection was not provided in the paper. FP–false positive.

[2] There are other ways to categorize catastrophic faults. For example, in Refs. [50,246], catastrophic faults are grouped into line-to-line, ground, and arc faults.

layer ANN model to classify line-line faults from irradiation, temperature, current, voltage, and power of the PV system with a 98.53% classification accuracy. Two ANN models were built in [230], one of which was for fault detection and the other one was for diagnosis. Both models took irradiation, temperature, current, and voltage as inputs, while the detection model output the binary conditions (i.e., faulty or healthy state) and the diagnosis output the fault type.

Besides ANN models, other AI algorithms are also applied in FD classification methods. For example, an SVM was used in [233] to learn the influence of blocking diode short-circuits on the PV power output, which achieved up to 75.8% accuracy. A stacked two-stage SVM classifier was used to efficiently detect line-line faults in a PV system under different operating conditions with minimum data [234]. A DT model was used in [235] to classify fault types with an accuracy of up to 99.98%. The DT achieved 99% in identifying fault types by using irradiation, temperature ambient, and power ratio. [237] classified four types of faults, i.e., line-line faults, degradation, open circuit, and partial shading, in a PV system using an RF model that only considered current and voltage with more than 98% accuracy. kNN is also adopted in PV FD. [238] applied kNN in detecting and classifying open circuit faults, line-line faults, partial shading with and without bypass diode faults, and partial shading, by using irradiation, temperature, current, and voltage, with a 98.70% accuracy.

Deep learning techniques are popularly employed in FD. For example, [239] used LSTM as feature extractors to construct features from current, voltage, and power for fault detection. The classification accuracies of normal condition, line-to-line fault, and hot spot fault were 99.23%, 98.78%, and 97.66%, respectively. A deep residual network (ResNet) was developed in [240] to classify 8 types of faults with a 95.778% accuracy. [241] transformed PV electrical time series data into graphs, which were used in CNN models for fault classification. The accuracy of the developed method was over 99%.

Most image-based FD methods utilize deep learning architectures, which is promising for large-scale detections. However, this special group of methods require sensors with high cost and sometimes cannot monitor the PV system continuously. For example, a series of work focused on detecting visible defects in PV modules, include dust-shading, encapsulant delamination, gridline corrosion, snail trails, and yellowing [242–244]. Aerial images were collected by an unmanned aerial vehicle (UAV) and labeled thereafter, which were used to train CNN models. The developed optimal CNN architecture achieved over 97% accuracy. In [245], more advanced images, namely, the infrared aerial images, were collected by a UAV, which could detect invisible defects.

4.3. Summary

In this section, AI-based PV array detection and PV fault detection are extensively reviewed. In both applications, deep learning techniques show promising and leading potentials in terms of accuracy and scalability, which will still be the future trend. The research in these fields could be further advanced by improvements from the following perspectives:

- Datasets. Compared to similar work, such as forecasting, publicly available datasets for detection in solar energy field are limited. There is only one open-source dataset for PV array detection, only covering 4 cities in California. The dataset for PV fault detection is even rare. This is mainly due to the significantly more effort required to collect, process, and label the dataset.
- Deep learning. Image-based detection is possibly one of the areas that could benefit the most from deep learning techniques. As computer vision is one of the most promising deep learning application fields, both PV array detection and PV fault detection could be further improved by employing popular deep learning architectures and methods.
- Application value. The application value of the two research topics, especially the PV array detection, is underestimated. It will be interesting to further explore the utilization of PV installation information in power system real-time operations.

5. Design optimization

Optimization is a promising approach worthy of study that aims to improve the system operation and enhance the economic benefit for PV systems. Compared with gradient-based and heuristic approaches, the metaheuristic algorithm is a high-level problem-independent stochastic method, which has been extensively employed to solve highly nonlinear and multi-modal problems under various complex constraints. Note that there are still debates on whether the metaheuristic algorithm belongs to AI, and it is noticed that several review papers directly include metaheuristic learning as one form of AI algorithms [45,43,48]. There also exist several review papers, covering a broader topic of hybrid energy systems and optimization methodologies [247,248]. In this review paper, we focus on AI-based optimization that is applied to PV systems.

5.1. Sizing optimization

The PV sizing problem mainly emphasizes the power balance in a stand-alone hybrid system that is composed of various kinds of energy sources, such as PV, wind, battery, diesel, and fuel cell. The sizing optimization is to optimize the size of different energy components to achieve the goal of reducing the cost and improving the supply reliability. Several typical PV sizing optimizations with heuristic algorithms are summarized in Table 11, which includes critical information of the optimization problems. Since most optimization approaches follow a standard optimization procedure (i.e., problem formulation, solver, and solution) [249], this section reviews AI-based PV system design optimizations from the perspectives of objective, design variable, constraint, and optimizer.

5.1.1. Objective function and design variable

Optimization goals are represented by objective functions, and several objectives in PV system optimization have been used in the literature: (i) the total cost that consist of initial investment, operation cost, and maintenance cost of the whole hybrid system, (ii) reliability metrics, such as the loss of power supply probability (LPSP), the loss of load probability (LOLP), and the loss of energy expectation (LOEE), as defined and highlighted in Table 12, iii) environmental benefits, such as the fuel consumption, CO_2 emission [255], and renewable factor. For a single-objective optimization problem, the total cost is usually selected as the objective function, in which two indices are optimized: the total system cost (TSC) of the estimated entire lifetime and the total annual cost (TAC). For example, the TSC in [251] consisted of the equipment cost, load curtailment cost, manufacturing cost of microgrids, and revenue of the power sale to the main grid; while [266] utilized TAC that consisted of the annual capital cost and annual maintenance cost as the optimization objective.

In some circumstances, it is insufficient to only consider the economic performance. Therefore, a multi-objective optimization approach considering the economic performance, reliability, and environmental impacts is of great significance to yield a balanced optimal solution. A multi-dimensional Pareto frontier of the targeted objectives can be generated, in which various optimal solutions with different emphases and weight factors can be selected based on different conditions. For instance, [254] considered both the TSC and LPSP for the PV-wind-battery system, which yielded a Pareto frontier as the final output. [267] extended the framework to several seasonal weather conditions, and generated several Pareto frontiers with different scenarios. [265] took into account the TAC, LPSP, and renewable factor as the objectives for a hybrid system, and solved the multi-objective optimization with a weighted sum method. [268] used a metaheuristic algorithm to

Table 11
Typical PV sizing optimizations using metaheuristic algorithm

Reference	Types	Decision variables	Objective (s)	Solver (s)	Highlights
[250]	• PV	• PV installation	• Total system cost	PSO	
	• Wind	• Wind turbine size			• Uncertainty and reliability evaluation
	• Battery	• Battery capacity			• Optimal installations of different energy sources
		• Sizing of other components			
[251]	• PV	• PV installation	• Total annual cost		
	• Battery	• Wind turbine size		PSO	• PSO is faster than GA
	• Fuel cell	• Fuel cell size		GA	• Hybrid system optimal sizing
	• Wind	• Sizes of other components			
[252]	• PV	• PV installation size	• Total system cost	Markov-based	• The Markov-based GA provided competitive computational cost
	• Diesel	• Diesel generator size			• Different constraint scenarios consist of LPSP and emission
	• Wind	• Wind turbine size			
[253]	• PV	• PV installation size	• Total annual costs	PSO/ IPSO/	• Eight metaheuristic algorithm were tested
	• Battery	• Battery capacity		GA/ ABSO/	• ABSO yielded more promising results
	• Wind	• Wind turbine size		TS/SA/ HS	
[254]	• PV	• PV module size	• Total system cost	GA	• Pareto frontier
	• Wind	• Wind turbine size	• Loss of power supply probability (LPSP)		
	• Battery	• Battery capacity			
		• Regulator, inverter			
[255]	• PV	• PV installation size	• Total system cost	MOPSO-	• Developed for mobile microgrid (ship)
	• Diesel	• Diesel generator size	• Diesel generator emission	NSGA-II	• PV/diesel/ battery is better than PV/ diesel
	• Battery	• Battery capacity			

determine the optimal number and type of renewable energy units by minimizing both the total system cost and the emissions of a stand-alone hybrid wind-PV-diesel energy system. To avoid unnecessary calculation of the entire Pareto frontier, as long as the weight factors of different objectives are empirically predefined, the multi-objective problem can be transformed into a single objective problem. (Note that if the sum of the weight factors equals to one, the solution is a subset of the original Pareto frontier.) For example, [269] considered the pollution of fuel emission as a social cost and combined it with the annual financial cost with different coefficients to optimize the real-time operation scheduling for a hybrid system.

Table 12
Representative PV sizing objective functions and constraints

Definition	Explanation and highlight	Reference
Loss of power supply probability (LPSP)	• The ratio of the power supply from a combined (PV and other renewable resources) system that is not able to supply the load	[256,257]
	• The most popular evaluation index	
Loss of load probability (LOLP)	• The ratio of total energy deficit to total load demand	[253]
	• Occurs when the demand surpasses the generation capacity	[258]
Loss of energy expectation (LOEE)	• The expected energy that has not been supplied	
	• Occurs when the generation capacity is not able to meet the hourly load demand	[259]
Expected energy not supplied (EENS)	• The expected energy that is not provided to the load	[260]
	•Occurs when the load demand exceeds the available generation capacity	[261]
		[262]
Equivalent loss factor (ELF)	• The ratio of effective load outage hours to the total number of hours	[263]
	• Contains information about both the number and magnitude of outages for a stand-alone network	
Deficiency of power supply probability (DPSP)	• The ratio of all deficiency of the power supply values to the overall load demand for a given period	
	• Highlights the insufficient power supply conditions	[264]
Renewable factor (RF)	• The ratio of renewable energy to the overall energy	
	• Effective in the cases with diesel generators or grid-connected networks	[265]

For design variables, it is seen from Table 11 that the majority of the reviewed papers have optimized sizing parameters of hybrid systems, such as PV installation capacity, PV panel area, battery capacity, wind turbine capacity, wind turbine swept area, and the sizing of other auxiliary subsystem or components [248]. The number of design variables highly depends on the assumptions and complexity of the studied case.

5.1.2. Constraint

In an optimization problem, constraints are retained to guarantee feasible solutions based on prerequisites and natural characteristics. For metaheuristic algorithms, it is necessary to identify the feasibility of every attempt before calculation to save computational cost. The most widely-employed constraints for the PV sizing problem are based on the power balance, design variable boundaries, economic limitations [270], reliability, voltage profile limitations, and environmental impacts [252], as summarized in Table 12. For instance, [266] introduced a constraint of LPSP to minimize the TAC, and the optimal results showed that the LPSP was satisfied in its lower boundary, indicating reasonable reliability performance. [271] minimized the cost of energy and the LOLE by putting a maximum constraint of the LOLE to guarantee the reliability and stability of the hybrid system. Similarly, [272] included the TAC, LOEE, and LOLE in the objective function with a reliability constraint of ELF. Results indicated the costs of the system depended on the reliability and the outage probability of major components.

It is worth mentioning that the aforementioned objectives and constraints are interchangeable in some cases according to the decision-making process. For instance, the reliability index LPSP in [273] is considered in the objective together with the energy cost, while LPSP is treated as a constraint with LLP in [256] to minimize the TSA.

5.1.3. Optimizer

Because of the complexity and non-convexity of the objective functions and constraints, metaheuristic algorithms are widely employed for PV sizing optimization, while deterministic searching and gradient-based optimization methods are inefficient in certain circumstances. Metaheuristic algorithms attempt to explore the searching space to select the best or a sufficiently good and feasible solution among all possible candidates in a stochastic manner. Several methods have been proposed to classify existing metaheuristic algorithms with respect to their properties, e.g., local search or global search, single or population-based, nature-inspired or metaphor-based algorithms [274,275]. In this paper, we attempt to classify metaheuristic algorithms based on the approaches to update candidate solutions [276], as categorized in Fig. 6.

Local search metaheuristic algorithms achieve the optimal or near-optimal solution by searching, replacing, and iterating from a single current solution around its neighborhood. This type of algorithm is a trajectory-based method that starts with a complete solution and improves the current solution within the smallest local moves by employing either the steepest descent or first moving strategy, in which the best move or the first better solution is selected, respectively. Popular local search algorithms include the variable neighborhood search (VNS) that considers a sequence of moves, simulated annealing (SA) that selects the next solution randomly, and tabu search (TS) that deals with their neighborhood. Because of their limitation in computational efficiency, their application in PV sizing optimization are relatively less popular compared to other metaheuristic algorithms. For instance, [277] employed SA to minimize a hybrid system's total cost considering the PV size, wind installation capacity, and battery size. Results showed that SA obtained a better result than the response surface methodology. To accelerate the convergence, local search algorithms can be used in conjunction with other direct searching or metaheuristic methods. For example, [270] compared the performance of both the TS and SA algorithms for the optimal sizing problem of a renewable-based autonomous power system. To expedite the process, a hybrid SA-TS algorithm was proposed by using SA to find an initial region and using TS to further improve the optimal solution.

Constructive metaheuristic algorithms start with an empty solution and construct solutions from their constituent elements. By adding the best possible elements during each iteration using a greedy algorithm, follow by a local search step, this type of algorithm is more efficient than the aforementioned local search algorithms in some cases [278]. The greedy randomized adaptive search procedure (GRASP) and ant colony optimization (ACO) are two of the most common constructive metaheuristic algorithms. Only ACO is reported in the literature for PV sizing optimization. For example, [279] employed ACO to perform sizing optimizations for three renewable energy systems including a hybrid system, a solar standalone system, and a wind standalone system. Results showed the final optimal solution and computational efficiency were promising. [280] also utilized the ACO algorithm to perform PV-related sizing and performance analysis of a PV-wind-battery hybrid energy

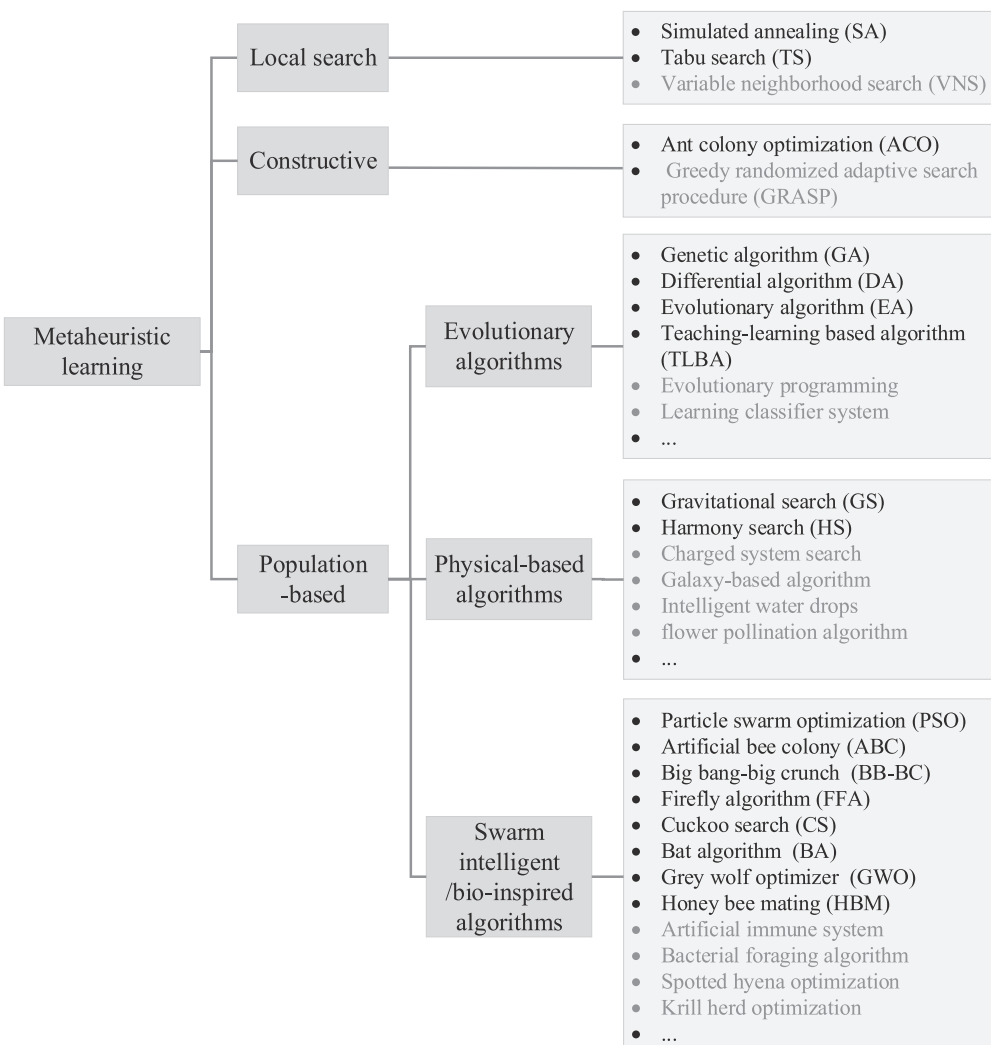

Fig. 6. Metaheuristic learning applied to PV system optimization (the most prevalent algorithms are in black).

system. The obtained optimal configuration was found to provide the minimal energy cost with excellent performance and reduced unmet load. Moreover, an improved ACO algorithm was proposed by [281] to shorten the computational time by selecting highlighted routes instead of all possible routes. The authors proved that decreasing the system cost and increasing the system reliability were conflicting with each other.

Population-based metaheuristic algorithms search near-optimal solutions by iteratively selecting and combining existing solutions using population characteristics. This type of algorithm is memory-based, which stores and compares solutions between the current step and the previous step. Except the algorithms discussed above, the rest heuristic algorithms can be categorized into this class. Popularly-used population-based metaheuristic algorithms include genetic algorithm (GA), particle swarm optimization (PSO), harmony search (HS), and cuckoo search (CS), [282], as summarized in Fig. 6.

Population-based algorithms have been widely employed in PV sizing optimization. For example, GA is an evolutionary algorithm that usually starts from a population of randomly generated individuals, followed by an iterative process of selection, crossover, and mutation until the solution converges. [283] employed GA to optimize the sizes of PV, wind, and battery, and found that a combination of different battery types is not favorable. [252] used GA to minimize the total cost while retaining the reliability with an LPSP constraint. The results showed that PV-wind hybrid systems featured lower system costs compared to PV-alone or wind-alone cases. [284] also used GA to obtain the optimal PV size in a grid-connected system. The study highlighted the effects of price fluctuation in a time-of-use tariff structure.

PSO is another nature-inspired well-known algorithm with fewer control parameters than GA, in which particles search through the problem space by following the current optimum particles and their own best historical solutions. For instance, [251] utilized PSO as an optimizer for the sizing of a PV-based hybrid system. [285] first optimized the size of a hybrid system with the standard PSO algorithm. Then to further evaluate the performance of PSO, several variants of PSO in terms of modification (MPSO), repulsion factor (PSO-RF), constriction factor (PSO-CF), and adaptive inertia weight (PSO-W), were compared and found that PSO-CF produced more promising results. [286] proposed a meta-PSO algorithm for sizing and results showed that the total cost of a hybrid wind-PV system was lower than that of a stand-alone wind or PV system.

Applications of other population-based metaheuristic algorithms are summarized as follows. The imperialist competitive algorithm (ICA) [259] was adopted to determine the optimal sizes of autonomous and non-autonomous hybrid green power system equipment by considering economics, reliability indices, and environmental emissions. It proved that purchasing grid-connected power resulted in a lower overall efficiency of the non-autonomous system. By modifying crossover and mutation operators of the crow search algorithm (CSA), an improved-CSA was developed to improve the load supply reliability by [287]. It was also reported that the improved-CSA performed better than both CSA and PSO. [288] proposed a new sizing approach based on the cuckoo search (CS) algorithm for grid-connected PV-wind-battery hybrid energy systems. It proved that CS had better accuracy, faster convergence, and less computation time compared to PSO.

Moreover, several specific hybrid algorithms were developed by combining two or more metaheuristic algorithms, which were mainly used to solve multi-objective optimization problems. For example, [262] utilized a hybrid big bang-big crunch theory (H-BB-BC) for optimal sizing of a stand-alone PV-wind-battery hybrid system. By taking advantage of the PSO capacities and a mutation operator, the algorithm was able to improve the exploration ability of the BB-BC algorithm and avoid trapping into local optimums. The results showed that the H-BB-BC algorithm was better than both the PSO and discrete harmony search (HS) algorithm. [255] used a multi-objective particle swarm optimization (MOPSO) algorithm combined with the non-dominated sorting genetic algorithm (NSGA-II) to solve the optimal sizing problem with respect to the investment cost and greenhouse gas emission. It revealed that the acquired net present cost of a hybrid PV-diesel-battery was less than that of a PV-diesel system.

It is important to note that although several comparative studies have been performed for metaheuristic algorithms regarding the convergence and computational efficiency, it is still insufficient to draw a safe conclusion for the ranking of various metaheuristic algorithms. The selection of a proper metaheuristic algorithm for the sizing problem of a PV-contained hybrid system highly depends on the assumptions and setting conditions.

5.2. Sizing and siting optimization

Compared to the aforementioned PV sizing, PV sizing and siting need to consider the topology impacts of renewable sources on a specific stand-alone or grid-connected energy system. In sequence, siting information is added to the design variable group, making the optimization problem more complicated and challenging. It is noticed that a large number of studies with metaheuristic algorithms have been performed in recent years, seeking to mitigate negative effects of high PV penetration or improve the economic efficiency by jointly optimizing the sizing and siting of distributed renewable energy systems [289]. This section is organized based on applications.

5.2.1. Pure PV

Due to the uncertain and variable characteristics of PV energy, only a small number of studies have been performed on sizing and siting optimization for systems that only have PV. For instance, [290] presented a GA-based approach to optimize the sizes and sites of solar panel units to minimize the power loss and improve the voltage profile. [291] determined the optimal penetration of PV units in the IEEE 37-bus network by using GA, while considering multiple buses integration under different scenarios. [292] presented a method to determine the size and placement of PV for a real 162-bus electric distribution network using GA. This study attempted to improve the current system by minimizing the total power loss and voltage deviation.

5.2.2. PV and battery

The PV-battery hybrid system is one of the most common settings, in which the battery is employed to mitigate the power unbalance between PV generation and load profile. For example, [293] proposed to use a unified PSO algorithm to determine the optimal sizing and siting of a battery storage system based on different PV penetration scenarios in the IEEE 33-bus system. [294] used a GA-based bi-level optimization method to reduce voltage fluctuations caused by high PV penetration, via optimizing the capacity and installation locations of a PV-battery energy system in a grid-connected power network. The approach yielded consistent solutions, and the achieved optimal sizing and siting were validated with the IEEE 8500-node test feeder.

5.2.3. Other hybrid system

Besides battery systems, PV also combines with other energy resources like wind, fuel cell, and diesel generators. For instance, [295] optimized the sizes, sites, and schedules for an off-grid hybrid system by using a two-staged PSO algorithm. The comparisons among three scenarios, i.e., diesel-only generator, diesel-PV, and diesel-PV-battery, revealed that the diesel-PV-battery system yielded the lowest cost while retaining voltage constraints. [296] presented a multi-objective optimization for the sizing and siting of a PV-wind-fuel cell hybrid energy system in a 70-bus system. An improved honey bee mating optimization (HBMO) algorithm was utilized and compared with GA, PSO, and TS algorithm. Results showed that proper siting and sizing are important for improving the voltage profile, reduce costs, emission, and losses of a distribution system. [297] presented a new hybrid grey wolf optimizer (GWO)-PSO technique to optimally allocate the size of PV and other distributed resources in order to reduce power losses. The

approach yielded better performance compared to conventional PSO and GWO, and was validated in IEEE 33-bus and 69-bus systems and a 78-bus real distribution system.

5.2.4. PV and electric vehicles

Moreover, the sizing and siting of electrical vehicles (EVs) charging stations have also been considered together with PV and other renewable energy resources. The EV charging load has a different profile compared to the conventional load profiles. An appropriate sizing, siting, and scheduling setting play an important role in system reliability and economic efficiency. For example, [298] proposed a GA-PSO-based optimal planning strategy, using varying renewable energy sources (PV-wind) to meet additional EV charging demands. The approach aimed to reduce power losses, voltage fluctuations, charging and demand supplying costs, and EV battery costs. Case studies on the IEEE 33-bus system proved the effectiveness of using GA-PSO to determine the optimal siting and sizing of PV, wind turbine, and EV charging stations simultaneously. Similarly, a multi-objective optimization was performed by [299] in 69-bus and 94-bus microgrids and solved using differential evolution (DE) algorithm. Results showed that both the optimal sizing and siting of PV-wind-EV charging station, as well as the optimal EVs charging schedules, helped improve the system performance.

5.3. Summary

In this section, the sizing and siting optimization for PV or PV-related hybrid systems are comprehensively reviewed. Research trends regarding PV sizing and siting could be briefly summarized as follows:

- Sizing. The sizing optimization usually involves with other types of renewable energy sources in a standalone system or grid-connected network if electricity power exchange is allowed. Both economic and reliability objectives need to considered during the optimization. Convergence and computational efficiency of different metaheuristic algorithms are highly dependent on detailed system conditions and parameters.
- Siting. The siting problem needs to be evaluated specifically based on different networks. Generally, a more complex network will enable a more complicated optimization problem. These challenges require more powerful solvers to reduce the computational cost, which affects the wide applications of metaheuristic algorithms.

Two aspects need to be improved in future studies.

- EVs. The rapid growth of EVs may raise new challenges for the sizing and siting problem in a hybrid energy system. The research on PV sizing and siting needs to be further enhanced by considering a larger power network with more distributed energy resources.
- Hybrid metaheuristic algorithms. Large-scale power networks tend to be employed in the sizing and siting optimization, which requires more powerful solvers with fast convergence and high accuracy. Though, the continuing improvements in modified and hybrid metaheuristic algorithms would benefit PV applications.

6. Optimal control

PV-relative control is a wide topic that involves PV power generation system and the controls of its relative devices and equipments like inverter and transformer. The Maximum power point tracking (MPPT) is one of the most important control process in PV systems. Extensive studies have found that the performance and efficiency of PV modules are highly affected by irradiation, ambient temperature, and load profiles [300]. To enhance the energy harvesting efficiency, it is important to accurately achieve MPPT under both normal uniform irradiation conditions and partial shading conditions. A large number of algorithms have been developed for MPPT, such as incremental conductance (INC), perturb and observe (P&O), open-circuit voltage method, and look up table method [301]. This paper mainly reviews the AI-based MPPT algorithms.

6.1. Fuzzy logic control

One of the most significant advantages of fuzzy logic control (FLC) is that the mathematical model of the system is not required, and the decision is based on an estimated value using approximate reasoning with the linguistic rules in the form of IF-THEN statements. This offers significant flexibility with respect to uncertainty and nonlinearity, making FLC relatively easier to be implemented in practice. For example, [302] employed the fuzzy logic algorithm to control a DC-DC converter in a stand-alone PV water pumping system under variable temperatures and insolation conditions, to improve the PV energy production efficiency. [303] established a fuzzy logic controller for MPPT in a PV system and optimized fuzzy membership functions and duty cycles using four different methods, including the teaching–learning based optimization (TLBO), firefly algorithm (FFA), biogeography based optimization (BBO), and PSO. Results showed that TLBO and FFA performed much better in terms of the MPPT convergence and tracking accuracy. Moreover, FLC predefined settings have a significant impact on the control performance. To mitigate the impacts of parameter tuning, [304] proposed a novel FLC by adding an intermediate variable β, aiming to simplify the fuzzy rule membership functions to cover wider operating conditions. It was observed that the converging speed in transient condition was improved and oscillations around the maximum power point were eliminated by employing this new approach.

6.2. Artificial neural network

ANN is mainly employed to identify energy systems and load profiles by establishing numerical black-box models. Note that, unlike traditional statistical approaches, ANN is able to model linear or non-linear systems without any implicit assumptions, which provides incomparable flexibility. For example, ANN was used to determine parameters of an emulated MPP locus and then embed them into a digital MPPT system [305]. It was noticed that the ANN model exhibited advantages like low computational requirement, fast-tracking speed, and high efficiency. To further explore the capability of ANN modeling, a comparative study between ANN and conventional incremental conductance method in grid-connected PV systems was conducted by [306], in which ANN's hyper-parameters were optimized using GA. The comparison showed that the improved ANN yielded preferable results. Different ANN-based hybrid algorithms have also been developed to estimate the maximum power point and perform accurate control. For instance, [307] employed ANN to estimate voltages as an input signal to an FLC controller, in which significant improvements of precision and tracking speed were observed. Besides system identifications, an NN-based controller of MPPT is another application. For example, [307] established an ANN-based MPPT controller in two steps, in which the first step was to optimize neural network parameters off-line while the second step was to use the optimal ANN-MPPT controller for MPPT online. Both experimental and simulation results showed that the ANN controller outperformed the P&O method in terms of both convergence and oscillations.

6.3. Particle swarm optimization

As discussed in Section 5, PSO is a population-based metaheuristic algorithm that has a decent convergence speed towards large non-convex problems. It is found from the literature that traditional PSO-based MPPT techniques may suffer unexpected power oscillations in a steady-state [308]. Consequently, several modifications of the conventional PSO were reported to improve the MPPT capability. For instance,

[309] conducted a modification for PSO on the initial value selection in MPPT optimization. Compared with a conventional hill-climbing method, the modified PSO under both uniform and partial shading conditions performed significantly better, which was also validated via a built-in hardware prototype. [310] improved the traditional PSO by dispersing particles and the algorithm was used to evaluate the variant of the global maximum power. The authors claimed that the improved PSO together with a shading pattern change reinitialization methodology performed the best in tracking the dynamic global maximum power. Similarly, Farh et al. (2019) modified the position and velocity vector of PSO to make it be able to follow the dynamic global maximum power under time-invariant partial shading conditions. The authors compared the MPPT performance among the improved PSO, a hybrid PSO-FLC approach (Farh et al., 2018), and a deep recurrent neural network (DRNN) (Farh et al., 2019). It was reported that the improved PSO algorithm was not as good as the other two algorithms in terms of steady-state oscillation, tracking speed, and accuracy, while the comparisons between the hybrid PSO-FLC and DRNN were not reported.

6.4. Ant colony optimization

By applying the ACO algorithm, the original MPPT problem is transformed into a probabilistic problem to mimic the social behavior of ants searching for a food source by finding the best path on a weighted graph. The ants marked walked routs with pheromones, and the concentration of pheromone may indicate the shortest route between their nest and food. ACO has been widely used for MPPT. For instance, [311] developed a proportional integral (PI)-based fractional open circuit voltage technique to enhance the ACO method for the MPPT problem in a stand-alone PV system. The proposed method outperformed traditional method in terms of satisfactory dynamic and steady-state performance. [312] proposed a hybrid searching method that integrated ACO in initial stages of MPP tracking due to its global searching ability and P&O for its local searching ability. Results showed that the hybrid approach performed better than the standard ACO. [313] updated the pheromone of the standard PSO algorithm with a random distribution search technique. The obtained results were compared with other algorithms including standard ACO, ANN, FLC, and PSO, and it proved that the PSO with modified pheromone had the best performance under variable atmospheric conditions.

6.5. Genetic algorithm

GA is a population-based metaheuristic algorithm based on evolutionary biological behavior. GA treats randomly generated initial candidate solutions as fixed-length chromosomes, and then iteratively searches the optimal or near-optimal solution via mutation, crossover, and selection. Due to its relatively slow speed of convergence, GA is rarely directly employed for PV MPPT. However, due to its advantages of global searching in which GA explores the search space using its various kinds of crossover methods, GA has incomparable virtues in tuning parameters for PI-based and FLC-based controllers. For instance, [314] utilized GA to tune and obtain the optimum parameters of a PI-based MPPT controller for a stand-alone PV-diesel hybrid energy system. Simulation results showed the effectiveness of the proposed controller in terms of dynamic response, voltage oscillation damping, and maximum overshooting. [315] optimized and implemented a GA-aided FLC controller for MPPT in a stand-alone PV system. The obtained results showed good tracking efficiency and rapid response to changes in environmental parameters. Moreover, GA can also be applied to train and tune hyper-parameters in an ANN model. [316] proposed a GA-optimized ANN-based MPPT approach, and results showed that the proposed approach satisfied the requirement of the optimum power supply and attenuated the fluctuations around the operating point.

6.6. Other metaheuristic algorithms

Besides the popular AI-based algorithms reviewed above, there are several other metaheuristic algorithms employed to address the MPPT problem. For example, an improved differential evolution (DE) algorithm was adopted by [317] to establish a single-ended primary-inductor converter, the feasibility of which was validated through physical implementation and experimentation. A firefly optimization algorithm (FOA) was proposed by [318] for the MPPT problem under partially shaded conditions. Compared with the traditional P&O and standard PSO method, both the experimental and simulation results presented a higher tracking efficiency and tracking speed of FOA. Cuckoo search (CS) is another well-known metaheuristic algorithm that exhibits advantages like fast convergence and high efficiency without expertise-based parameter tuning. CS was employed by [319] for MPPT, the results showed that CS outperformed both P&O and PSO regarding tracking capability, transient behavior, and convergence.

6.7. Summary

In this section, AI-based algorithms show promising control performance for PV MPPT. It is seen that modified and hybrid algorithms exhibit distinguished advantages over single algorithms with respect to searching the global maximum power or avoiding the oscillation around MPP under partial shading conditions. Due to differences in parameter settings and evaluation criteria, it is challenging to accurately rank the MPPT capabilities among different algorithms. A comprehensive comparison among all these algorithms based on the same PV models and shading conditions is still needed to provide a better understanding.

7. Other topics

In addition to the above four main application themes, ML is also widely used in solar integration from other perspectives, such as radiation modeling, synthetic solar data generation, solar energy assessment, etc. Table 13 lists some representative work in these applications.

Solar radiation modeling is to quantify the amount of incident solar radiation from non-radiation parameters (estimation), radiation parameters (decomposition), and the same radiation parameters with different settings (transposition). Solar radiation modeling is not a new topic. However, different from application like solar agricultural meteorology and architectural simulations, solar radiation modeling for solar power integration requires a much higher temporal resolution (at least hourly resolution). Therefore, AI-based solar radiation modeling research is relatively less than other similar work, such as solar forecasting. In this research theme, satellite imagery channels, weather parameters, calendar parameters, and cloud coverage are the most popular inputs. ANN, RF, and GBM are the prevalent ML algorithms. For example, ANN models were used to correlate satellite data in [322,323] to estimate different components of irradiance or radiance. [321] developed an RF model to estimate the global solar radiation from sunshine ratio, humidity, ambient temperature, and several calendar features. [327] developed XGBoost separation models to estimate DNI and DIF from GHI and weather parameters. [320] explored the use of ensemble ML models, i.e., RF and GBM, to model the daily and hourly solar irradiance from calendar features and clearness index.

Synthetic solar data generation is to generate solar irradiance or power data by spatio-temporal correlation or downscaling. For example, [329] developed ANN models to downscale the daily solar radiation into hourly solar radiation. The ANN models generated hourly solar radiance data from daily solar radiation, hour angle, and sunset hour angle with a 3.1% RMSE. [330] used a generalized linear model that included non-Gaussian mixtures to synthesize 1 min GHI data. It was found that the simulated data showed good coverage properties and temporal correlation structures. A similar method was reported in [331], which used the Gaussian mixture model to downscale 30 min GHI into 1 min GHI.

Table 13
Other AI-based solar integration applications.

Reference	Application	Method	Input	Output
[320]	Radiation estimation	RF, GBM	Clearness index, solar time, day number	GHI, DNI, DHI, DIF
[321]	Radiation estimation	RF	Sunshine ratio, weather and calendar features	Global solar radiation
[322]	Radiation estimation	ANN	Satellite images	DNI
[323]	Radiation estimation	ANN ensemble	Satellite images, calendar features	DNI, DHI, GHI
[324]	Radiation estimation	ANN	Weather features, local time	GHI
[325]	Radiation estimation	GBM	Weather features, cloud cover, visibility	DIF
[156]	Radiation estimation	ANN, SVR	Reflectivity, clear sky radiation, cloud index	Global solar radiation
[326]	Radiation estimation	ANN	Weather features	Global solar radiation
[327]	Radiation decomposition	XGBoost	GHI and weather parameters	DNI, DIF
[328]	Radiation transposition	ANN, SVR	Radiation on the horizontal surface	Radiation on the tilted surface
[329]	Synthetic data generation	ANN	daily radiation, hour angle, sunset hour angle	Hourly radiation
[330]	Synthetic data generation	Generalized linear model	30 min GHI	1 min GHI
[331]	Synthetic data generation	Gaussian mixture model	30 min GHI	1 min GHI
[332]	Synthetic data generation	Gaussian copula	Spatio-temporal features	Irradiance
[333]	Resource assessment	SVR	Building and population features	Roof characteristics
[334]	Resource assessment	SVM	Roof characteristics	Roof shape type
[335]	Resource assessment	RF	Building characterstics, weather features	Rooftop PV installation potential

Spatio-temporal features were used in [332] by a Gaussian copula based on the propagating cloud field. The method was able to downscale hourly solar irradiance data to higher resolutions in both space and time.

ML models were also used in solar resource assessment. In [333], SVR models were used to estimate the roof characteristics, including available roof area for PV installation, shading factors from neighbouring buildings and trees on the building roofs, global tilted solar radiation on non-horizontal surfaces, from urban features. Then the geographical potential and the technical potential of distributed PV systems in Switzerland were quantified by modeling the PV electricity generation from the estimated building characteristics. [334] used SVM models to classify solar roof-shape of 10,085 buildings in the city of Geneva in Switzerland. The model successfully identifies six types of roof shapes based on their useful area for PV installations and the potential for receiving solar energy. The same research group combined the previous two studies by using RF models to estimate the solar and urban variables in entire Switzerland and found that the rooftop PV production can cover 25 of Switzerland's demand in 2017 [335].

AI techniques, with their powerful learning capabilities, have been applied to many other research themes that are related to solar energy integration, such as PV tracking system control, cloud estimation, cloud type and weather type classification, and microgrid management. The review of these broad topics is beyond the scope of this paper.

8. Discussion

AI has been identified as a key driver to facilitate solar power penetration. An ever-increasing body of solar AI literature motivates a systematic and comprehensive way of literature review in this field. It will be especially beneficial to junior scholars who need an overall picture of the field, conference organizers who desire to offer mini-tracks and workshops with emerging topics, and journal editors who want to document the history and develop particular streams of research [24]. This paper performs a solar AI literature review by combining text mining and human expertise. Text mining plays a vital and irreplaceable role in collecting, categorizing, and ranking related literature. However, at the current stage, text mining is still a complementary approach. Therefore, this paper relies on human expertise to summarize, analyze, and compare solar AI papers.

With the development of sensing and smart meter technologies, a large volume of data is collected from different sources for solar AI research. It is important to identify diverse and optimal data according to the research task. For example, NWP should be incorporated in longer-term solar forecasting, while sky images may be only used in intra-hour solar forecasting. Feature engineering should be performed if it is necessary. To facilitate reproducibility of the solar AI research, it is suggested to open-source the data or at least use publicly available datasets in case studies. Some popular open-source datasets for forecasting and detection are listed in Table 14. A number of tools are also available to get access to these datasets, such as the OpenSolar [184] and SolarData [336] packages. Since the AI models are getting more complex, it is essential to conduct experiments with enough data. For example, at least one year of data should be used to test forecasting models. Researchers are encouraged to compare their work with the state-of-the-art benchmarks and publish their codes along with the paper.

From the algorithm/model perspective, parameter/hyperparameter optimization is necessary to optimize models. The optimization process should be described and visualized to understand the modeling process. For example, in CNN, how many CNN layers should be included in each CNN block, and how many CNN blocks should be used to construct the model? The model training process should also be reported to justify the successful learning. AI/ML models are highly affected by the data.

Table 14
Open-source time series datasets for BTM modeling and forecasting.

Name/ Reference	Parameters	Spatial coverage	Resolution/ length	Field
Microgen [337]	PV power, meta data	7,000 + locations	30 min/1 year	Forecasting
Solcast [338]	PV power, meta data	1,287 locations	10 min/7 months	Forecasting
UCSD [339]	Meteorological variables, NWP	1 location	1 min/3 years	Forecasting
NSRDB [340]	Meteorological variables	US	30 min/25 + years	Forecasting
NREL/SRRL [341]	Meteorological variables	1 location	1 min/40 years	Forecasting
SUFRAD [342]	Meteorological variables	7 locations	1 or 3 min/ 25 years	Forecasting
Duke PV Imagery [343]	Satellite aerial images	US	30 cm/NA	Detection

Therefore, the same data processing, including pre-processing and post-processing, should be ensured when comparing different models. For complex methodologies that consist of a sequence of techniques, improvements from each technique should be analyzed. For example, some research achieves forecasting improvements by decomposing time series into multiple signals and predicting each signal with an advanced ML model. It is important to identify which step leads to the improvement and how much improvement each step contributes.

In this review paper, AI techniques in four main domains are reviewed. Specifically, in solar forecasting, the application of deep learning is still at its early stage, compared to shallow ML. The development of deep learning techniques is expected to capture spatial and temporal patterns in multiple data sources in solar forecasting data in a more efficient way. Deep learning-based solar forecasting could advance or at least be competitive with shallow ML methods. Among different types of data, image-based solar forecasting is still less researched. Inclusion and optimization of input combinations from various sources are expected to further improve AI-based forecasting performance. Compared to load and wind forecasting, solar forecasting in probabilistic forms is also lagging behind. Probabilistic solar forecasting will help power system operators better manage uncertainties associated with forecasts, therefore, becoming an emerging topic and will be continuously under investigation.

In the solar PV detection theme, two major fields that rely heavily on AI techniques are solar PV array detection and fault detection. Both fields are critical to solar PV integration and operations. Similar to solar forecasting, deep learning is becoming prevalent in this field and will be the focus in the future. However, there is a lack of consent to using standard datasets and metrics for case studies. Any research that facilitates the open-source research in this field will receive high attention. Regarding the dataset, power system measurements, such as the Advanced Metering Infrastructure (AMI) or higher-level data, and the non-intrusive measurements, such as satellite aerial images, will be especially beneficial. In terms of metrics, popular classification evaluation metrics, such as confusion matrix-based metrics and ROC curve-based metrics should be used. In addition, the application value of this field is not fully recognized. This should be enhanced by combining detection tasks with power system operations.

Regarding the siting and sizing of PV system, hybrid metaheuristic algorithms should be a foreseeable trend as the penetration of PV and EVs increase continually in some local grid, making the grid network more complicated. Another potential trend is to optimize the sizing and siting of PV and other distributed generation sources to establish reconfigurable microgrids and provide stronger resilience and reliability against extreme weather conditions and other disruptive events.

For the MPPT techniques, hybrid approaches that combine multiple AI-based control methods may play a dominating role in future studies, due to its excellent advantages like superior tracking control and global optimization capability compared to traditional control algorithms. Moreover, more efforts are also expected to alleviate its side effects including high computational complexity and poor real-time control ability.

9. Conclusion

This paper conducted a comprehensive taxonomical review on artificial intelligence (AI) applications in solar photovoltaic (PV) system grid integration. The bibliographic infrastructure was constructed by 2,772 papers that were collected, analyzed, and categorized using text mining techniques. Four main solar AI research themes were identified by the latest review papers, term frequencies, and the latent Dirichlet allocation (LDA) method, which were forecasting, detection, design optimization, and control. All the papers were assigned to one of the research themes, where a total of 330 papers that are most relevant and popular were reviewed. It is important to note that the text mining results may miss certain papers due to the keywords selection. The review paper focuses on recent AI applications to solar photovoltaic system forecasting, detection, optimization, and control. Other AI-related solar grid integration topics, such as dispatching scheduling, reactive power control, protection coordination, are beyond the scope this review.

Solar forecasting, including irradiance and power output forecasting, is the theme that publishes the most papers. Recent forecasting papers advanced forecasting techniques from feature engineering, model optimization, and post-processing steps. Several emerging topics, such as image-based forecasting, deep learning forecasting, and probabilistic forecasting, were identified by the latest forecasting papers. Specifically, automated feature selection and construction should be conducted using advanced optimization methods, which incorporates data from various sources, especially sky images, shadow images, and satellite images. The development of deep learning techniques is and will still be a hot topic in solar forecasting. Probabilistic solar forecasting will be a promising field to mitigate solar forecasting uncertainties.

AI-based detection detection consists of PV array detection and PV fault detection. PV fault detection used either AI regression methods or classification methods, both of which showed accurate results. PV array detection is a new research field, where deep learning contributed to most papers. However, AI-based PV detection was less popular compared to solar forecasting. This was largely due to the relatively lack of publicly available datasets and underestimated application value in power system operations.

Metaheuristic learning is extensively used in sizing and sizing-siting optimization problems in PV-contained systems, which was comprehensively reviewed by algorithms and by different systems. It was found that the sizing optimization usually combined PV with other kinds of renewable energy resources in standalone or grid-connected systems. The sizing-siting optimization was far more complicated, which required more powerful solvers to reduce the computational cost. The PV-electrical vehicle optimization was an emerging scenario that was less researched. In these optimization problems, economic efficiency and reliability were the two prior objectives, and hybrid metaheuristic learning was the trend. At last, a comprehensive comparison of all the state-of-the-art algorithms is of great necessity to fully evaluate the large collection of metaheuristic learning methods.

ANN and metaheuristic learning applications in PV maximum power point control were also extensively reviewed. The modified and hybrid algorithms exhibited distinguished advantages over single algorithms in terms of searching the global maximum power and avoiding the oscillation around the maximum power point under partial shading conditions.

CRediT authorship contribution statement

Cong Feng: Data curation, Formal analysis, Investigation, Methodology, Validation, Writing - original draft. **Yuanzhi Liu:** Data curation, Formal analysis, Investigation, Methodology, Validation, Writing - original draft. **Jie Zhang:** Conceptualization, Investigation, Methodology, Project administration, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Sorrell Steve. Reducing energy demand: A review of issues, challenges and approaches. Renew Sustain Energy Rev 2015;47:74–82.
[2] IRENA. Renewable power generation costs in 2018. Technical report, International Renwable Energy Agency, Abu Dhabi, 2019.
[3] Timilsina Govinda R, Kurdgelashvili Lado, Narbel Patrick A. Solar energy: Markets, economics and policies. Ren Sustainable Energy Rev 2012;16:449–65.

[4] USDOE. Solar Forecasting [Online]. Available: https://www.energy.gov/eere/solar/improving-accuracy-solar-forecasting-funding-opportunity, 2012.
[5] USDOE. Solar Forecasting 2 [Online]. Available: https://www.energy.gov/eere/solar/solar-forecasting-2, 2017.
[6] ARPA-E. Arpa-e funding opportunity announcements [Online], 2019.
[7] NSF. Energy, power, control, and networks [Online]. Available: https://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505249&org=ECCS&from=home, 2019.
[8] NSF. Smart power protection devices for photovoltaic installations [Online]. Available. 2013.
[9] NSF. Image modeling and machine learning algorithms for utility-scale solar panel monitoring [Online]. Available: https://www.nsf.gov/awardsearch/showAward?AWD_ID=1646542, 2016.
[10] NSF. Data-driven voltage var optimization enabling extreme integration of distributed solar energy [Online]. Available: https://www.nsf.gov/awardsearch/showAward?AWD_ID=1929975&HistoricalAwards=false, 2019.
[11] EWeLiNE. Eweline: Development of innovative weather and power forecast models for the grid integration of weather dependent energy sources [Online]. Available: http://www.projekt-eweline.de/en/index.html, 2019.
[12] Andreas Ulbig, Daniel Cajoos, Michel Gasche, Marc Laeser, Stephan Koch, René Hoffmann, et al. Gridsense sologrid pilot project—using decentralized artificial intelligence for making distribution grids resilient. In: CIRED Workshop 2016, IET, 2016, p. 1–3.
[13] PrognoNetz. Prognonetz – an intelligent ampacity forecast for overhead lines [Online]. Available: https://www.itiv.kit.edu/english/6518.php, 2019.
[14] IRENA. Artificial intelligence and big data. Technical report, International Renwable Energy Agency, Abu Dhabi, 2019.
[15] He Zhuo, Zhang Yan, Li Huiyuan. Self-inspection cleaning device for photovoltaic power plant based on machine vision. In: IOP Conference Series: Earth and Environmental Science, vol. 242. IOP Publishing; 2019. p. 032020.
[16] Kumar Nallapaneni Manoj, Sudhakar K, Samykano M, Jayaseelan V. On the technologies empowering drones for intelligent monitoring of solar photovoltaic power plants. Procedia Comput Sci 2018;133:585–93.
[17] de Silva D, Pang Z, Osipov Evgeny, Vyatkin Valeriy. Guest editorial: Special section on developments in artificial intelligence for industrial informatics. IEEE Trans Industr Inf 2019;15(6):3690–2.
[18] Ramos Carlos, Liu Chen-Ching. Ai in power systems and energy markets. IEEE Intell Syst 2011;26(2):5–8.
[19] Xu Zhao, Zhao Junhua. Solar energy harvesting, storage and utilization [Online]. Available. 2018.
[20] Yang Dazhi, Gueymard Christian A, Kleissl Jan. Submission of data article is now open. Sol Energy 2018;171(27):A1–2.
[21] Carlos FM Coimbra. Looking ahead with the journal of renewable and sustainable energy: Volume 11 and beyond, 2019.
[22] O'Mara-Eves Alison, Thomas James, McNaught John, Miwa Makoto, Ananiadou Sophia. Using text mining for study identification in systematic reviews: a systematic review of current approaches. Systematic Rev 2015;4(1):5.
[23] Yang Dazhi, Kleissl Jan, Gueymard Christian A, Pedro Hugo TC, Coimbra Carlos FM. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. Sol Energy 2018;168:60–101.
[24] Delen Dursun, Crossland Martin D. Seeding the survey and analysis of research literature with text mining. Expert Syst Appl 2008;34:1707–20.
[25] Moro Sérgio, Cortez Paulo, Rita Paulo. Business intelligence in banking: A literature analysis from 2002 to 2013 using text mining and latent dirichlet allocation. Expert Syst Appl 2015;42(3):1314–24.
[26] Korhonen Anna, Séaghdha Diarmuid O, Silins Ilona, Sun Lin, Högberg Johan, Stenius Ulla. Text mining for literature review and knowledge discovery in cancer risk assessment and research. PloS One 2012;7(4):e33427.
[27] Abbe Adeline, Grouin Cyril, Zweigenbaum Pierre, Falissard Bruno. Text mining applications in psychiatry: a systematic literature review. Int J Methods Psychiatric Res 2016;25(2):86–100.
[28] Kwartler Ted. Text mining in practice with R. John Wiley & Sons; 2017.
[29] Falagas Matthew E, Pitsouni Eleni I, Malietzis George A, Pappas Georgios. Comparison of pubmed, scopus, web of science, and google scholar: strengths and weaknesses. The FASEB J 2008;22(2):338–42.
[30] Mellit Adel, Kalogirou Soteris A, Drif Mahmoud. Application of neural networks and genetic algorithms for sizing of photovoltaic systems. Renewable Energy 2010;35(12):2881–93.
[31] Ansari Md Fahim, Chatterji S, Iqbal Atif. A fuzzy logic control scheme for a solar photovoltaic system for a maximum power point tracker. Int J Sustain Energ 2010;29(4):245–55.
[32] Caputo Davide, Grimaccia Francesco, Mussetta Marco, Zich Riccardo E. Photovoltaic plants predictive model by means of ann trained by a hybrid evolutionary algorithm. In: The 2010 International joint conference on neural networks (IJCNN). IEEE; 2010. p. 1–6.
[33] Sun Yuchi, Szűcs Gergely, Brandt Adam R. Solar pv output prediction from video streams using convolutional neural networks. Energy Environ Sci 2018;11(7):1811–8.
[34] Medhat Elsayed, Melike Erol-Kantarci, Burak Kantarci, Lei Wu, and Jie Li. Low-latency communications for community resilience microgrids: A reinforcement learning approach. IEEE Trans Smart Grid, 2019.
[35] Yu Jiafan, Wang Zhecheng, Majumdar Arun, Rajagopal Ram. Deepsolar: A machine learning framework to efficiently construct a solar deployment database in the united states. Joule 2018;2(12):2605–17.
[36] Voyant Cyril, Notton Gilles, Kalogirou Soteris, Nivet Marie-Laure, Paoli Christophe, Motte Fabrice, Fouilloy Alexis. Machine learning methods for solar radiation forecasting: A review. Renewable Energy 2017;105:569–82.
[37] Atika Qazi, Fayaz H, Wadi A, Ram Gopal Raj, Rahim NA, Waleed Ahmed Khan. The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review. J Cleaner Prod 2015;104:1–12.
[38] Zendehboudi Alireza, Baseer MA, Saidur R. Application of support vector machine models for forecasting solar and wind energy resources: A review. J Cleaner Prod 2018;199:272–85.
[39] Obando Edgar Darío, Carvajal Sandra Ximena, Agudelo Jairo Pineda. Solar radiation prediction using machine learning techniques: A review. IEEE Latin America Trans 2019;17(04):684–97.
[40] Wang Huaizhi, Lei Zhenxing, Zhang Xian, Zhou Bin, Peng Jianchun. A review of deep learning for renewable energy forecasting. Energy Convers Manage 2019;198:111799.
[41] Jesús Ferrero Bermejo, Juan F Gómez Fernández, Fernando Olivencia Polo, and Adolfo Crespo Márquez. A review of the use of artificial neural network models for energy and reliability prediction. a study of the solar pv, hydraulic and wind energy sources. ApplSci, 2019;9(9):1844.
[42] Muhammad Naveed Akhter, Saad Mekhilef, Hazlie Mokhlis, and Noraisyah Mohamed Shah. Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. IET Renewable Power Generat 2019;13(7): 1009–1023.
[43] Seyedmahmoudian M, Horan B, Kok Soon T, Rahmani R, Muang Than Oo A, Mekhilef S, Stojcevski A. State of the art artificial intelligence-based mppt techniques for mitigating partial shading effects on pv systems–a review. Renew Sustain Energy Rev 2016;64:435–55.
[44] Karabacak Kerim, Cetin Numan. Artificial neural networks for controlling wind–pv power systems: A review. Renew Sustain Energy Rev 2014;29:804–27.
[45] Youssef Ayman, El-Telbany Mohammed, Zekry Abdelhalim. The role of artificial intelligence in photo-voltaic systems design and control: A review. Renew Sustain Energy Rev 2017;78:72–9.
[46] Adel Mellit and Soteris A Kalogirou. Mppt-based artificial intelligence techniques for photovoltaic systems and its implementation into field programmable gate array chips: Review of current status and future perspectives. Energy, 2014;70:1–21.
[47] Elobaid Lina M, Abdelsalam Ahmed K, Zakzouk Ezeldin E. Artificial neural network-based photovoltaic maximum power point tracking techniques: a survey. IET Renew Power Gener 2015;9(8):1043–63.
[48] Zahraee SM, Khalaji Assadi M, Saidur R. Application of artificial intelligence methods for hybrid energy system optimization. Renew Sustain Energy Rev 2016;66:617–30.
[49] Triki-Lahiani Asma, Abdelghani Afef Bennani-Ben, Slama-Belkhodja Ilhem. Fault detection and monitoring systems for photovoltaic installations: A review. Renew Sustain Energy Rev 2018;82:2680–92.
[50] Mellit Adel, Tina Giuseppe Marco, Kalogirou Soteris A. Fault detection and diagnosis methods for photovoltaic systems: A review. Renew Sustain Energy Rev 2018;91:1–17.
[51] Pillai Dhanup S, Rajasekar N. Metaheuristic algorithms for pv parameter identification: A comprehensive review with an application to threshold setting for fault detection in pv systems. Renew Sustain Energy Rev 2018;82:3503–25.
[52] Rodrigues Eugénio, Gomes Álvaro, Gaspar Adélio Rodrigues, Antunes Carlos Henggeler. Estimation of renewable energy and built environment-related variables using neural networks–a review. Renew Sustain Energy Rev 2018;94:959–88.
[53] Elsheikh Ammar H, Sharshir Swellam W, Elaziz Mohamed Abd, Kabeel AE, Guilan Wang, Haiou Zhang. Modeling of solar energy systems using artificial neural network: A comprehensive review. Sol Energy 2019;180:622–39.
[54] Mosavi Amir, Salimi Mohsen, Ardabili Sina Faizollahzadeh, Rabczuk Timon. Shahaboddin Shamshirband, and Annamaria Varkonyi-Koczy. State of the art of machine learning models in energy systems, a systematic review. Energies 2019;12(7):1301.
[55] Kow Ken Weng, Wong Yee Wan, Rajkumar Rajparthiban Kumar, Rajkumar Rajprasad Kumar. A review on performance of artificial intelligence and conventional method in mitigating pv grid-tied related power quality events. Renew Sustain Energy Rev 2016;56:334–46.
[56] Arun Rajkumar, Venkatasubramaniyan Suresh CE, Madhavan Veni, Narasimha Murthy MN. On finding the natural number of topics with latent dirichlet allocation: Some observations. In: Pacific-Asia conference on knowledge discovery and data mining. Springer; 2010. p. 391–402.
[57] Cao Juan, Xia Tian, Li Jintao, Zhang Yongdong, Tang Sheng. A density-based method for adaptive lda model selection. Neurocomputing 2009;72(7–9):1775–81.
[58] Deveaud Romain, SanJuan Eric, Bellot Patrice. Accurate and effective latent concept modeling for ad hoc information retrieval. Document numérique 2014;17(1):61–84.
[59] Thomas L Griffiths and Mark Steyvers. Finding scientific topics. Proc National Acad Sci, 2004;101(suppl 1):5228–5235.
[60] Ali Pourmousavi S, Cifala AS, Nehrir MH. Impact of high penetration of pv generation on frequency and voltage in a distribution feeder. In: 2012 North American Power Symposium (NAPS). IEEE; 2012. p. 1–8.
[61] Cui Mingjian, Zhang Jie, Feng Cong, Florita Anthony R. Yuanzhang Sun, and Bri-Mathias Hodge. Characterizing and analyzing ramping events in wind power, solar power, load, and netload. Renewable Energy 2017;111:227–44.

[62] Kazemi Mehdi, Siano Pierluigi, Sarno Debora, Goudarzi Arman. Evaluating the impact of sub-hourly unit commitment method on spinning reserve in presence of intermittent generators. Energy 2016;113:338–54.
[63] Jie Zhang, Bri-Mathias Hodge, Siyuan Lu, Hendrik F Hamann, Brad Lehman, Joseph Simmons, Edwin Campos, Venkat Banunarayanan, Jon Black, and John Tedesco. Baseline and target values for regional and point pv power forecasts: Toward improved solar forecasting. Solar Energy, 2015;122:804–819.
[64] Lori Bird, Jaquelin Cochran, and Xi Wang. Wind and solar energy curtailment: Experience and practices in the united states. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States), 2014.
[65] Zahedi Ahmad. Maximizing solar pv energy penetration using energy storage technology. Renew Sustain Energy Rev 2011;15(1):866–70.
[66] Cui Mingjian, Zhang Jie, Hodge Bri-Mathias, Lu Siyuan, Hamann Hendrik F. A methodology for quantifying reliability benefits from improved solar power forecasting in multi-timescale power system operations. IEEE Trans Smart Grid 2017;9(6):6897–908.
[67] Cui Mingjian, Zhang Jie. Estimating ramping requirements with solar-friendly flexible ramping product in multi-timescale power system operations. Appl Energy 2018;225:27–41.
[68] CAISO. Business practice manual for market operations [Online]. 2019.
[69] Antonanzas Javier, Osorio Natalia, Escobar Rodrigo, Urraca Ruben. Francisco J Martinez-de Pison, and Fernando Antonanzas-Torres. Review of photovoltaic power forecasting. Sol Energy 2016;136:78–111.
[70] Mathieu David F, Ramahatana Pierre-Julien Trombe, Lauret Philippe. Probabilistic forecasting of the solar irradiance with recursive arma and garch models. Sol Energy 2016;133:55–72.
[71] Li Yanting, Su Yan, Shu Lianjie. An armax model for forecasting the power output of a grid connected photovoltaic system. Renewable Energy 2014;66:78–89.
[72] Dazhi Yang, Stefano Alessandrini, Javier Antonanzas, Fernando Antonanzas-Torres, Viorel Badescu, Hans Georg Beyer, et al. Verification of deterministic solar forecasts. Sol Energy, 2020.
[73] Lauret Philippe, David Mathieu, Pinson Pierre. Verification of solar irradiance probabilistic forecasts. Sol Energy 2019;194:254–71.
[74] Majidpour Mostafa, Nazaripouya Hamidreza, Chu Peter, Pota Hemanshu R, Gadh Rajit. Fast univariate time series prediction of solar power for real-time control of energy storage system. Forecasting 2019;1(1):107–20.
[75] David Mathieu, Luis Mazorra Aguiar, Lauret Philippe. Comparison of intraday probabilistic forecasting of solar irradiance using only endogenous data. Int J Forecast 2018;34(3):529–47.
[76] Wu Ji, Chan Chee Keong, Yu Zhang, Xiong Bin Yu, Zhang Qing Hai. Prediction of solar radiation with genetic approach combing multi-model framework. Renewable Energy 2014;66:132–9.
[77] Pedro Hugo TC, Coimbra Carlos FM. Assessment of forecasting techniques for solar power production with no exogenous inputs. Sol Energy 2012;86(7): 2017–28.
[78] Al-Messabi Naji, Li Yun, El-Amin Ibrahim, Goh Cindy. Forecasting of photovoltaic power yield using dynamic neural networks. In: The 2012 International Joint Conference on Neural Networks (IJCNN). IEEE; 2012. p. 1–5.
[79] Voyant Cyril, Muselli Marc, Paoli Christophe, Nivet Marie-Laure. Optimization of an artificial neural network dedicated to the multivariate forecasting of daily global radiation. Energy 2011;36(1):348–59.
[80] Vagropoulos Stylianos I, Chouliaras GI, Kardakos Evaggelos G, Simoglou Christos K, Bakirtzis Anastasios G. Comparison of sarimax, sarima, modified sarima and ann-based models for short-term pv generation forecasting. In: 2016 IEEE International Energy Conference (ENERGYCON). IEEE; 2016. p. 1–6.
[81] Bouzgou Hassen, Gueymard Christian A. Minimum redundancy–maximum relevance with extreme learning machines for global solar radiation forecasting: Toward an optimized dimensionality reduction for solar time series. Sol Energy 2017;158:595–609.
[82] Voyant Cyril, Randimbivololona Prisca, Nivet Marie Laure, Paoli Christophe, Muselli Marc. Twenty four hours ahead global irradiation forecasting using multi-layer perceptron. Meteorol Appl 2014;21(3):644–55.
[83] Lauret Philippe, Voyant Cyril, Soubdhan Ted, David Mathieu, Poggi Philippe. A benchmarking of machine learning techniques for solar radiation forecasting in an insular context. Sol Energy 2015;112:446–57.
[84] Martí Pau, Gasque María. Improvement of temperature-based ANN models for solar radiation estimation through exogenous data assistance. Energy Convers Manage Feb 2011;52(2):990–1003.
[85] Maria Grazia De Giorgi, Maria Malvoni, and Paolo Maria Congedo. Photovoltaic power forecasting using statistical methods: impact of weather data. IET Sci, Measur Technol 2014;8(3):90–97.
[86] Russo M, Leotta G, Pugliatti PM, Gigliucci G. Genetic programming for photovoltaic plant output forecasting. Sol Energy 2014;105:264–73.
[87] Rana Mashud, Koprinska Irena, Agelidis Vassilios G. Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. Energy Convers Manage 2016;121:380–90.
[88] Persson Caroline, Bacher Peder, Shiga Takahiro, Madsen Henrik. Multi-site solar power forecasting using gradient boosted regression trees. Sol Energy 2017;150: 423–36.
[89] Voyant Cyril, Muselli Marc, Paoli Christophe, Nivet Marie-Laure. Numerical weather prediction (nwp) and hybrid arma/ann model to predict global radiation. Energy 2012;39(1):341–55.
[90] Mazorra Aguiar L, Pereira B, Lauret P, Díaz F, David M. Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. Renewable Energy Nov 2016;97:599–610.
[91] Qing Xiangyun, Niu Yugang. Hourly day-ahead solar irradiance prediction using weather forecasts by lstm. Energy 2018;148:461–8.
[92] Yang Dazhi. A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically-based, ensemble, and skill (ropes). J Renewable Sustainable Energy 2019;11(2):022701.
[93] Han Seung Jang, Kuk Yeol Bae, Hong-Shik Park, and Dan Keun Sung. Solar power prediction based on satellite images and support vector machine. IEEE Trans n Sustainable Energy 2016;7(3):255–1263.
[94] Marquez Ricardo, Pedro Hugo TC, Coimbra Carlos FM. Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to anns. Sol Energy 2013;92:176–88.
[95] Dong Zibo, Yang Dazhi, Reindl Thomas, Walsh Wilfred M. Satellite image analysis and a hybrid esss/ann model to forecast solar irradiance in the tropics. Energy Convers Manage 2014;79:66–73.
[96] Davy Robert J, Huang Jing R, Troccoli Alberto. Improving the accuracy of hourly satellite-derived solar irradiance by combining with dynamically downscaled estimates using generalised additive models. Sol Energy Oct 2016;135:854–63.
[97] Marquez Ricardo, Gueorguiev Vesselin G, Coimbra Carlos FM. Forecasting of global horizontal irradiance using sky cover indices. J Sol Energy Eng 2013;135 (1):011017.
[98] Marquez Ricardo, Coimbra Carlos FM. Intra-hour dni forecasting based on cloud tracking image analysis. Sol Energy 2013;91:327–36.
[99] Chu Yinghao, Pedro Hugo TC, Coimbra Carlos FM. Hybrid intra-hour dni forecasts with sky image processing enhanced by stochastic learning. Sol.Energy 2013;98:592–603.
[100] Chu Yinghao, Urquhart Bryan, Gohari Seyyed MI, Pedro Hugo TC, Kleissl Jan, Coimbra Carlos FM. Short-term reforecasting of power output from a 48 mwe solar pv plant. Sol Energy 2015;112:68–77.
[101] Chu Yinghao, Pedro Hugo TC, Li Mengying, Coimbra Carlos FM. Real-time forecasting of solar irradiance ramps with smart image processing. Sol Energy 2015;114:91–104.
[102] Pedro Hugo TC, Coimbra Carlos FM. Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances. Renewable Energy 2015;80:770–82.
[103] Feng Cong, Cui Mingjian, Hodge Bri-Mathias, Siyuan Lu, Hamann Hendrik F, Zhang Jie. Unsupervised clustering-based short-term solar forecasting. IEEE Trans Sustainable Energy Oct 2019;10(4):2174–85.
[104] Feng Cong, Zhang Jie. Hourly-similarity based solar forecasting using multi-model machine learning blending. In: 2018 IEEE Power & Energy Society General Meeting (PESGM). IEEE; 2018. p. 1–5.
[105] Feng Cong, Cui Mingjian, Lee Meredith, Zhang Jie, Hodge Bri-Mathias, Lu Siyuan, Hamann Hendrik F. Short-term global horizontal irradiance forecasting based on sky imaging and pattern recognition. In: 2017 IEEE Power & Energy Society General Meeting. IEEE; 2017. p. 1–5.
[106] Crisosto Cristian, Hofmann Martin, Mubarak Riyad, Seckmeyer Gunther. One-hour prediction of the global solar irradiance from all-sky images using artificial neural networks. Energies Oct 2018;11(11):2906.
[107] Scolari Enrica, Sossan Fabrizio, Haure-Touzé Mathia, Paolone Mario. Local estimation of the global horizontal irradiance using an all-sky camera. Sol Energy Oct 2018;173:1225–35.
[108] Dinesh Pothineni, Martin R Oswald, Jan Poland, and Marc Pollefeys. Kloudnet: Deep learning for sky image analysis and irradiance forecasting. In: German Conference on Pattern Recognition, Springer, 2018, p. 535–51.
[109] Zhao Xin, Wei Haikun, Wang Hai, Zhu Tingting, Zhang Kanjian. 3d-cnn-based feature extraction of ground-based cloud images for direct normal irradiance prediction. Sol. Eergy 2019;181:510–8.
[110] Feng Cong, Zhang Jie. Solarnet: A deep convolutional neural network for solar forecasting via sky images. In: 2020 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). IEEE; 2020. p. 1–5.
[111] Sun Yuchi, Venugopal Vignesh, Brandt Adam R. Short-term solar power forecast with deep learning: Exploring optimal input and output configuration. Sol Energy Aug 2019;188:730–41.
[112] Venugopal Vignesh, Sun Yuchi, Brandt Adam R. Short-term solar PV forecasting using computer vision: The search for optimal CNN architectures for incorporating sky images and PV generation history. J Renewable Sustainable Energy Nov 2019;11(6):066102.
[113] Vaz AGR, Elsinga B, Van Sark WGJHM, Brito MC. An artificial neural network to assess the impact of neighbouring photovoltaic systems in power forecasting in utrecht, the netherlands. Renewable Energy 2016;85:631–41.
[114] Boudewijn Elsinga and Wilfried GJHM van Sark. Short-term peer-to-peer solar forecasting in a network of photovoltaic systems. Appl Energy 2017;206: 1464–1483.
[115] Huang Chao, Wang Long, Lai Loi Lei. Data-driven short-term solar irradiance forecasting based on information of neighboring sites. IEEE Trans Industr Electron 2018;66(12):9918–27.
[116] Gutierrez-Corea Federico-Vladimir, Manso-Callejo Miguel-Angel, Moreno-Regidor Maria-Pilar, Manrique-Sancho Maria-Teresa. Forecasting short-term solar irradiance based on artificial neural networks and data from neighboring meteorological stations. Sol.Energy 2016;134:119–31.
[117] Zagouras Athanassios, Pedro Hugo TC, Coimbra Carlos FM. On the role of lagged exogenous variables and spatio–temporal correlations in improving the accuracy of solar forecasting methods. Renewable Energy 2015;78:203–18.
[118] Huang Jing, Perry Matthew. A semi-empirical approach using gradient boosting and k-nearest neighbors regression for gefcom2014 probabilistic solar power forecasting. Int J Forecast 2016;32(3):1081–6.

[119] Zhaoxuan Li, SM Rahman, Rolando Vega, and Bing Dong. A hierarchical approach using machine learning methods in solar photovoltaic energy production forecasting. Energies 2016;9(1):55.
[120] Long Huan, Zhang Zijun, Yan Su. Analysis of daily solar power prediction with data-driven approaches. Appl Energy 2014;126:29–37.
[121] Pedro Hugo TC, Coimbra Carlos FM. Short-term irradiance forecastability for various solar micro-climates. Sol Energy 2015;122:587–602.
[122] Zhang Yue, Beaudin Marc, Taheri Raouf, Zareipour Hamidreza, Wood David. Day-ahead power output forecasting for small-scale solar photovoltaic electricity generators. IEEE Trans Smart Grid 2015;6(5):2253–62.
[123] Liu Zhao, Zhang Ziang. Solar forecasting by k-nearest neighbors method with weather classification and physical model. In: 2016 North American Power Symposium (NAPS). IEEE; 2016. p. 1–6.
[124] Li Yanting, He Yong, Su Yan, Shu Lianjie. Forecasting the daily power output of a grid-connected photovoltaic system based on multivariate adaptive regression splines. Appl Energy 2016;180:392–401.
[125] Azhar Ahmed Mohammed and Zeyar Aung. Ensemble learning approach for probabilistic forecasting of solar power generation. Energies 2016;9(12):1017.
[126] Urraca R, Antonanzas J, Alia-Martinez M, Martinez-de Pison FJ, Antonanzas-Torres F. Smart baseline models for solar irradiation forecasting. Energy Convers Manage 2016;108:539–48.
[127] Ferlito S, Adinolfi G, Graditi G. Comparative analysis of data-driven methods online and offline trained to the forecasting of grid-connected photovoltaic plant production. Appl Energy 2017;205:116–29.
[128] Chu Yinghao, Coimbra Carlos FM. Short-term probabilistic forecasts for direct normal irradiance. Renewable Energy 2017;101:526–36.
[129] Gigoni Lorenzo, Betti Alessandro, Crisostomi Emanuele, Franco Alessandro, Tucci Mauro, Bizzarri Fabrizio, Mucci Debora. Day-ahead hourly forecasting of power generation from photovoltaic plants. IEEE Trans Sustainable Energy 2017; 9(2):831–42.
[130] Chao-Rong Chen and Unit Three Kartini. k-nearest neighbor neural network models for very short-term global solar irradiance forecasting based on meteorological data. Energies 20174;10(2):186.
[131] Alfadda Abdullah, Rahman Saifur, Pipattanasomporn Manisa. Solar irradiance forecast using aerosols measurements: A data driven approach. Sol Energy 2018; 170:924–39.
[132] Pedro Hugo TC, Coimbra Carlos FM, David Mathieu, Lauret Philippe. Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts. Renewable Energy 2018;123:191–203.
[133] Benmouiza Khalil, Cheknane Ali. Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models. Energy Convers Manage 2013;75:561–9.
[134] Chen SX, Gooi HB, Wang MQ. Solar radiation forecast based on fuzzy logic and neural networks. Renewable Energy 2013;60:195–201.
[135] Koca Ahmet, Oztop Hakan F, Varol Yasin, Koca Gonca Ozmen. Estimation of solar radiation using artificial neural networks with different input parameters for mediterranean region of anatolia in turkey. Expert Syst Appl 2011;38(7): 8756–62.
[136] Liu Luyao, Liu Diran, Sun Qie, Li Hailong, Wennersten Ronald. Forecasting power output of photovoltaic system using a bp network method. Energy Procedia 2017; 142:780–6.
[137] Ding Ming, Wang Lei, Bi Rui. An ann-based approach for forecasting the power output of photovoltaic system. Procedia Environ Sci 2011;11:1308–15.
[138] Sara Pereira, Paulo Canhoto, Rui Salgado, Maria João Costa. Development of an ann based corrective algorithm of the operational ecmwf global horizontal irradiation forecasts. Sol Energy 2019;185:387–405.
[139] Klingler Anna-Lena, Teichtmann Lukas. Impacts of a forecast-based operation strategy for grid-connected pv storage systems on profitability and the energy system. Sol Energy 2017;158:861–8.
[140] Paoli Christophe, Voyant Cyril, Muselli Marc, Nivet Marie-Laure. Forecasting of preprocessed daily solar radiation time series using neural networks. Sol Energy 2010;84(12):2146–60.
[141] Romero Andrés F, Quilumba Franklin L, Arcos Hugo N. Short-term active power forecasting of a photovoltaic power plant using an artificial neural network. In: 2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM). IEEE; 2017. p. 1–5.
[142] Wang Fei, Mi Zengqiang, Shi Su, Zhao Hongshan. Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters. Energies 2012;5(5):1355–70.
[143] Amrouche Badia, Le Pivert Xavier. Artificial neural network based daily local forecasting for global solar radiation. Appl Energy 2014;130:333–41.
[144] Adel Mellit, Alessandro Massi Pavan. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected pv plant at trieste, italy. Sol Energy 2010;84(5):807–821.
[145] Voyant Cyril, Notton Gilles, Darras Christophe, Fouilloy Alexis, Motte Fabrice. Uncertainties in global radiation time series forecasting using machine learning: The multilayer perceptron case. Energy 2017;125:248–57.
[146] Lago Jesus, De Brabandere Karel, De Ridder Fjo, De Schutter Bart. Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data. Sol Energy Oct 2018;173:566–77.
[147] Benghanem Mohamed, Mellit Adel. Radial basis function network-based prediction of global solar radiation data: application for sizing of a stand-alone photovoltaic system at al-madinah, saudi arabia. Energy 2010;35(9):3751–62.
[148] Chen Changsong, Duan Shanxu, Cai Tao, Liu Bangyin. Online 24-h solar power forecasting based on weather type classification using artificial neural network. Sol Energy 2011;85(11):2856–70.
[149] Yadav Amit Kumar, Sharma Vikrant, Malik Hasmat, Chandel SS. Daily array yield prediction of grid-interactive photovoltaic plant using relief attribute evaluator based radial basis function neural network. Renew Sustain Energy Rev 2018;81: 2115–27.
[150] Hossain Monowar, Mekhilef Saad, Danesh Malihe, Olatomiwa Lanre, Shamshirband Shahaboddin. Application of extreme learning machine for short term output power forecasting of three grid-connected pv systems. J Cleaner Prod 2017;167:395–405.
[151] Behera Manoja Kumar, Majumder Irani, Nayak Niranjan. Solar photovoltaic power forecasting using optimized modified extreme learning machine technique. Eng Sci Technol Int J 2018;21(3):428–38.
[152] Majumder Irani, Dash PK, Bisoi Ranjeeta. Variational mode decomposition based low rank robust kernel extreme learning machine for solar irradiation forecasting. Energy Convers Manage 2018;171:787–806.
[153] Tang Pingzhou, Chen Di, Hou Yushuo. Entropy method combined with extreme learning machine method for the short-term photovoltaic power generation forecasting. Chaos Solitons Fractals 2016;89:243–8.
[154] Bouzgou Hassen. A fast and accurate model for forecasting wind speed and solar radiation time series based on extreme learning machines and principal components analysis. J Renewable Sustainable Energy 2014;6(1):013114.
[155] Sameer Al-Dahidi, Osama Ayadi, Jehad Adeeb, Mohammad Alrbai, Bashar R Qawasmeh. Extreme learning machines for solar photovoltaic power predictions. Energies 2018;11(10):2725.
[156] Cornejo-Bueno L, Casanova-Mateo C, Sanz-Justo J, Salcedo-Sanz S. Machine learning regressors for solar radiation estimation from satellite data. Sol Energy 2019;183:768–75.
[157] Ni Qiang, Zhuang Shengxian, Sheng Hanming, Kang Gaoqiang, Xiao Jian. An ensemble prediction intervals approach for short-term pv power forecasting. Sol Energy 2017;155:1072–83.
[158] Han Yutong, Wang Ningbo, Ma Ming, Zhou Hai, Dai Songyuan, Zhu Honglu. A pv power interval forecasting based on seasonal model and nonparametric estimation algorithm. Sol Energy 2019;184:515–26.
[159] Javier Huertas-Tato, Ricardo Aler, Inés M Galván, Francisco J Rodríguez-Benítez, Clara Arbizu-Barrena, David Pozo-Vázquez. A short-term solar radiation forecasting system for the iberian peninsula. part 2: Model blending approaches based on machine learning. Sol Energy 2020;195:685–696.
[160] Joao Gari da Silva Fonseca, Takashi Oozeki, Takumi Takashima, Gentarou Koshimizu, Yoshihisa Uchida, and Kazuhiko Ogimoto. Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in kitakyushu, japan. Progress in photovoltaics: Res Appl 2012;20(7):874–882.
[161] Zeng Jianwu, Qiao Wei. Short-term solar power prediction using a support vector machine. Renewable Energy 2013;52:118–27.
[162] Cheng Hsu-Yung, Yu Chih-Chang, Lin Sian-Jing. Bi-model short-term solar irradiance prediction using support vector regressors. Energy 2014;70:121–7.
[163] Dong Zibo, Yang Dazhi, Reindl Thomas, Walsh Wilfred M. A novel hybrid approach based on self-organizing maps, support vector regression and particle swarm optimization to forecast solar irradiance. Energy 2015;82:570–7.
[164] Li Jiaming, Ward John K, Tong Jingnan, Collins Lyle, Platt Glenn. Machine learning for solar irradiance forecasting of photovoltaic system. Renewable Energy 2016;90:542–53.
[165] Bae Kuk Yeol, Jang Han Seung, Sung Dan Keun. Hourly solar irradiance prediction based on support vector machine and its error analysis. IEEE Trans Power Syst 2016;32(2):935–45.
[166] Jiang He, Dong Yao. Forecast of hourly global horizontal irradiance based on structured kernel support vector machine: A case study of tibet area in china. Energy Convers Manage 2017;142:307–21.
[167] Sun Shaolong, Wang Shouyang, Zhang Guowei, Zheng Jiali. A decomposition-clustering-ensemble learning approach for solar radiation forecasting. Sol Energy 2018;163:189–99.
[168] Antonanzas J, Pozo-Vázquez D, Fernandez-Jimenez LA, Martinez-de Pison FJ. The value of day-ahead forecasting for photovoltaics in the spanish electricity market. Sol Energy 2017;158:140–6.
[169] Awad Mariette, Khanna Rahul. Support Vector Regression. Apress; 2015.
[170] Benali L, Notton G, Fouilloy A, Voyant C, Dizene R. Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. Renewable Energy 2019;132: 871–84.
[171] Fouilloy Alexis, Voyant Cyril, Notton Gilles, Motte Fabrice, Paoli Christophe, Nivet Marie-Laure, Guillot Emmanuel, Duchaud Jean-Laurent. Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability. Energy 2018;165:620–9.
[172] Torres-Barrán Alberto, Alonso Álvaro, Dorronsoro José R. Regression tree ensembles for wind energy and solar radiation prediction. Neurocomputing 2019; 326:151–60.
[173] Yagli Gokhan Mert, Yang Dazhi, Srinivasan Dipti. Automatic hourly solar forecasting using machine learning models. Renew Sustain Energy Rev 2019;105: 487–98.
[174] David John Gagne II, McGovern Amy, Haupt Sue Ellen, Williams John K. Evaluation of statistical learning configurations for gridded solar irradiance forecasting. Sol Energy 2017;150:383–93.
[175] Bessa Ricardo J, Trindade Artur, Silva Cátia SP, Miranda Vladimiro. Probabilistic solar power forecasting in smart grids using distributed information. Int J Electrical Power Energy Syst 2015;72:16–23.

[176] Verbois Hadrien, Rusydi Andrivo, Thiery Alexandre. Probabilistic forecasting of day-ahead solar irradiance using quantile gradient boosting. Sol Energy 2018; 173:313–27.
[177] Andrade José R, Bessa Ricardo J. Improving renewable energy forecasting with a grid of numerical weather predictions. IEEE Trans Sustainable Energy 2017;8(4): 1571–80.
[178] Wang Jidong, Li Peng, Ran Ran, Che Yanbo, Zhou Yue. A short-term photovoltaic power prediction model based on the gradient boost decision tree. Appl Sci 2018; 8(5):689.
[179] Raza Muhammad Qamar, Mithulananthan Nad, Summerfield Alex. Solar output power forecast using an ensemble framework with neural predictors and bayesian adaptive combination. Sol Energy 2018;166:226–41.
[180] Abuella Mohamed, Chowdhury Badrul. Random forest ensemble of support vector regression models for solar power forecasting. In: 2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT). IEEE; 2017. p. 1–5.
[181] Bessa Ricardo J, Trindade Artur, Miranda Vladimiro. Spatial-temporal solar power forecasting for smart grids. IEEE Trans Industr Inf 2014;11(1):232–41.
[182] Yagli Gokhan Mert, Yang Dazhi, Gandhi Oktoviano, Srinivasan Dipti. Can we justify producing univariate machine-learning forecasts with satellite-derived solar irradiance? Appl Energy 2020;259:114122.
[183] Liu Liping, Zhan Mengmeng, Bai Yang. A recursive ensemble model for forecasting the power output of photovoltaic systems. Sol Energy 2019;189: 291–8.
[184] Feng Cong, Yang Dazhi, Hodge Bri-Mathias, Zhang Jie. Opensolar: Promoting the openness and accessibility of diverse public solar datasets. Sol Energy 2019;188: 1369–79.
[185] Gábor I Nagy, Gergő Barta, Sándor Kazi, Gyula Borbély, and Gábor Simon. Gefcom 2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach. Int J Forecast 2016;32(3): 1087–1093.
[186] Hong Tao, Pinson Pierre, Fan Shu. Global energy forecasting competition 2012: 2014.
[187] Catur Hilman E, Tridianto TH Ariwibowo, Rohman Budiman PA. Forecasting of power output of 2-axis solar tracked pv systems using ensemble neural network. In: 2017 International Electronics Symposium on Engineering Technology and Applications (IES-ETA). IEEE; 2017. p. 152–6.
[188] Ren Ye, Suganthan PN, Srikanth N. Ensemble methods for wind and solar power forecasting—a state-of-the-art review. Renew Sustain Energy Rev 2015;50:82–91.
[189] Gensler André, Henze Janosch, Sick Bernhard, Raabe Nils. Deep learning for solar power forecasting—an approach using autoencoder and lstm neural networks. In: 2016 IEEE international conference on systems, man, and cybernetics (SMC). IEEE; 2016. p. 002858–65.
[190] Alzahrani Ahmad, Shamsi Pourya, Dagli Cihan, Ferdowsi Mehdi. Solar irradiance forecasting using deep neural networks. Procedia Comput Sci 2017;114:304–13.
[191] Zhang Jinsong, Verschae Rodrigo, Nobuhara Shohei, Lalonde Jean-François. Deep photovoltaic nowcasting. Sol Energy 2018;176:267–76.
[192] Husein Munir, Chung Il-Yop. Day-ahead solar irradiance forecasting for microgrids using a long short-term memory recurrent neural network: A deep learning approach. Energies 2019;12(10):1856.
[193] Abdel-Nasser Mohamed, Mahmoud Karar. Accurate photovoltaic power forecasting models using deep lstm-rnn. Neural Comput Appl 2019;31(7): 2727–40.
[194] Srivastava Shikhar, Lessmann Stefan. A comparative study of lstm neural networks in forecasting day-ahead global horizontal irradiance with satellite data. Sol Energy 2018;162:232–47.
[195] Wang Huaizhi, Yi Haiyan, Peng Jianchun, Wang Guibin, Liu Yitao, Jiang Hui, Liu Wenxin. Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. Energy Convers Manage 2017;153: 409–22.
[196] Feng Cong, Zhang Jie. Solarnet: A sky image-based deep convolutional neural network for intra-hour solar forecasting. Sol Energy 2020;204:71–8.
[197] Wang Fei, Yili Yu, Zhang Zhanyao, Li Jie, Zhen Zhao, Li Kangping. Wavelet decomposition and convolutional lstm networks based improved deep learning model for solar irradiance forecasting. Appl Sci 2018;8(8):1286.
[198] Lee Woonghee, Kim Keonwoo, Park Junsep, Kim Jinhee, Kim Younghoon. Forecasting solar power using long-short term memory and convolutional neural networks. IEEE Access 2018;6:73068–80.
[199] Wang Kejun, Qi Xiaoxia, Liu Hongda. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. Appl Energy 2019;251:113315.
[200] Talha Ahmad Siddiqui, Samarth Bharadwaj, Shivkumar Kalyanaraman. A deep learning approach to solar-irradiance forecasting in sky-videos. In: 2019 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, Jan 2019.
[201] Hossin Mohammad, Sulaiman MN. A review on evaluation metrics for data classification evaluations. Int J Data Mining Knowledge Manage Process 2015;5 (2):1.
[202] Fatourechi Mehrdad, Ward Rabab K, Mason Steven G, Huggins Jane, Schlögl Alois, Birch Gary E. Comparison of evaluation metrics in classification applications with imbalanced datasets. In: 2008 Seventh International Conference on Machine Learning and Applications2008 Seventh International Conference on Machine Learning and Applications. IEEE; 2008. p. 777–82.
[203] José Hernández-Orallo, Peter Flach, Cèsar Ferri. A unified view of performance metrics: translating threshold choice into expected classification loss. J Machine Learn Res 2012;13(Oct):2813–2869.
[204] Jordan M. Malof, Rui Hou, Leslie M. Collins, Kyle Bradbury, Richard Newell. Automatic solar photovoltaic panel detection in satellite imagery. In: 2015 International Conference on Renewable Energy Research and Applications (ICRERA). IEEE, Nov 2015.
[205] Jordan M. Malof, Kyle Bradbury, Leslie M. Collins, Richard G. Newell, Alexander Serrano, Hetian Wu, Sam Keene. Image features for pixel-wise detection of solar photovoltaic arrays in aerial imagery using a random forest classifier. In: 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA). IEEE, Nov 2016.
[206] Malof Jordan M, Bradbury Kyle, Collins Leslie M, Newell Richard G. Automatic detection of solar photovoltaic arrays in high resolution aerial imagery. Appl Energy 2016;183:229–40.
[207] Jordan M. Malof, Leslie M. Collins, Kyle Bradbury, Richard G. Newell. A deep convolutional neural network and a random forest classifier for solar photovoltaic array detection in aerial imagery. In: 2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA). IEEE, Nov 2016.
[208] Vladimir Golovko, Sergei Bezobrazov, Alexander Kroshchanka, Anatoliy Sachenko, Myroslav Komar, Andriy Karachka. Convolutional neural network based solar photovoltaic panel detection in satellite photos. In: 2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). IEEE, Sep 2017.
[209] Vladimir Golovko, Alexander Kroshchanka, Sergei Bezobrazov, Anatoliy Sachenko, Myroslav Komar, Oleksandr Novosad. Development of solar panels detector. In: 2018 International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T). IEEE, Oct 2018.
[210] Jordan M. Malof, Leslie M. Collins, Kyle Bradbury. A deep convolutional neural network, with pre-training, for solar photovoltaic array detection in aerial imagery. In: 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, Jul 2017.
[211] Jie Yongshi, Yue Anzhi, Liu Shunxi, Huang Qingqing, Yu Jingbo Chen, Meng Yupeng Deng, Yu Zongyang. Photovoltaic power station identification using refined encoder–decoder network with channel attention and chained residual dilated convolutions. J Appl Remote Sens Jan 2020;14(01):1.
[212] Joseph Camilo, Rui Wang, Leslie M. Collins, Kyle Bradbury, Jordan M. Malof. Application of a semantic segmentation convolutional neural network for accurate automatic detection and mapping of solar photovoltaic arrays in aerial imagery, 2018.
[213] Mainzer Kai, Killinger Sven, McKenna Russell, Fichtner Wolf. Assessment of rooftop photovoltaic potentials at the urban level using publicly available geodata and image recognition techniques. Sol Energy Oct 2017;155:561–73.
[214] Venkata Ramakrishna Padullaparthi, Venkatesh Sarangan, Anand Sivasubramaniam. suncover: Estimating the hidden behind-the-meter solar rooftop and battery capacities in grids. In: 2019 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), IEEE, 2019. p. 1–5.
[215] Killinger Sven, Lingfors David, Saint-Drenan Yves-Marie, Moraitis Panagiotis, van Sark Wilfried, Taylor Jamie, Engerer Nicholas A, Bright Jamie M. On the search for representative characteristics of PV systems: Data collection and analysis of PV system azimuth, tilt, capacity, yield and shading. Sol Energy 2018;173:1087–106.
[216] Harry Wirth and Karin Schneider. Recent facts about photovoltaics in Germany. Fraunhofer ISE, 92, 2015.
[217] Zhang Xiaochen, Grijalva Santiago. A data-driven approach for detection and estimation of residential PV installations. IEEE Trans Smart Grid 2016;7(5): 2477–85.
[218] Jordan M. Malof, Boning Li, Bohao Huang, Kyle Bradbury, Artem Stretslov. Mapping solar array location, size, and capacity using deep learning and overhead imagery, 2019.
[219] Rui Wang, Joseph Camilo, Leslie M. Collins, Kyle Bradbury, Jordan M. Malof. The poor generalization of deep convolutional networks to aerial imagery from new geographic locations: an empirical study with solar array detection. In: 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). IEEE, Oct 2017.
[220] Mohamed AH, Nassar AM. New algorithm for fault diagnosis of photovoltaic energy systems. Int J Comput Appl 2015;114(9).
[221] Chao Kuei-Hsiang, Chen Pi-Yun, Wang Meng-Hui, Chen Chao-Ting. An intelligent fault detection method of a photovoltaic module array using wireless sensor networks. Int J Distrib Sens Netw 2014;10(5):540147.
[222] Jiang Lian Lian, Maskell Douglas L. Automatic fault detection and diagnosis for photovoltaic systems using combined artificial neural network and analytical based methods. In: 2015 International Joint Conference on Neural Networks (IJCNN). IEEE; 2015. p. 1–8.
[223] Riley Daniel, Johnson Jay. Photovoltaic prognostics and heath management using learning algorithms. In: 2012 38th IEEE Photovoltaic Specialists Conference. IEEE; 2012. p. 001535–9.
[224] Leva Sonia, Mussetta Marco, Ogliari Emanuele. Pv module fault diagnosis based on microconverters and day-ahead forecast. IEEE Trans Industr Electron 2018;66 (5):3928–37.
[225] Mekki H, Mellit Adel, Salhi H. Artificial neural network-based modelling and fault detection of partial shaded photovoltaic modules. Simul Model Pract Theory 2016;67:1–13.
[226] Brofferio Sergio C, Antonini Alessio, Galimberti Gianluca, Galeri Dario. A method for estimating and monitoring the power generated by a photovoltaic module based on supervised adaptive neural networks. In: 2011 IEEE International Conference on Smart Measurements of Future Grids (SMFG) Proceedings. IEEE; 2011. p. 148–53.
[227] Karatepe Engin, Hiyama Takashi, et al. Controlling of artificial neural network for fault diagnosis of photovoltaic array. In: 2011 16th International Conference on Intelligent System Applications to Power Systems. IEEE; 2011. p. 1–6.
[228] Mohd Nafis Akram, Saeed Lotfifard. Modeling and health monitoring of dc side of photovoltaic array. IEEE Trans Sustainable Energy 2015;6(4):1245–1253.

[229] Dhimish Mahmoud, Holmes Violeta, Mehrdadi Bruce, Dales Mark. Comparing mamdani sugeno fuzzy logic and rbf ann network for pv fault detection. Renewable Energy 2018;117:257–74.
[230] Garoudja Elyes, Chouder Aissa, Kara Kamel, Silvestre Santiago. An enhanced machine learning based approach for failures detection and diagnosis of pv systems. Energy Convers Manage 2017;151:496–513.
[231] Chen Zhicong, Wu Lijun, Cheng Shuying, Lin Peijie, Wu Yue, Lin Wencheng. Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and iv characteristics. Appl Energy 2017;204:912–31.
[232] Yue Wu, Chen Zhicong, Lijun Wu, Lin Peijie, Cheng Shuying, Peimin Lu. An intelligent fault diagnosis approach for pv array based on sa-rbf kernel extreme learning machine. Energy Procedia 2017;105:1070–6.
[233] Wail Rezgui, Leïla-Hayet Mouss, Nadia Kinza Mouss, Mohamed Djamel Mouss, Mohamed Benbouzid. A smart algorithm for the diagnosis of short-circuit faults in a photovoltaic generator. In: 2014 First International Conference on Green Energy ICGE 2014, IEEE, 2014. p. 139–143.
[234] Yi Zhehan, Etemadi Amir H. Line-to-line fault detection for photovoltaic arrays based on multiresolution signal decomposition and two-stage support vector machine. IEEE Trans Industr Electron 2017;64(11):8546–56.
[235] Zhao Ye, Yang Ling, Lehman Brad, de Palma Jean-François, Mosesian Jerry, Lyons Robert. Decision tree-based fault detection and classification in solar photovoltaic arrays. In: 2012 Twenty-Seventh Annual IEEE Applied Power Electronics Conference and Exposition (APEC). IEEE; 2012. p. 93–9.
[236] Benkercha Rabah, Moulahoum Samir. Fault detection and diagnosis based on c4. 5 decision tree algorithm for grid connected pv system. Sol Energy 2018;173: 610–34.
[237] Chen Zhicong, Han Fuchang, Wu Lijun, Yu Jinling, Cheng Shuying, Lin Peijie, Chen Huihuang. Random forest based intelligent fault diagnosis for pv arrays using array voltage and string currents. Energy Convers Manage 2018;178: 250–64.
[238] Madeti Siva Ramakrishna, Singh SN. Modeling of pv system based on experimental data for fault detection using knn method. Sol Energy 2018;173: 139–51.
[239] Appiah Albert Yaw, Zhang Xinghua, Ayawli Ben Beklisi Kwame, Kyeremeh Frimpong. Long short-term memory networks based automatic feature extraction for photovoltaic array fault diagnosis. IEEE. Access 2019;7:30089–101.
[240] Chen Zhicong, Chen Yixiang, Wu Lijun, Cheng Shuying, Lin Peijie. Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions. Energy Convers Manage 2019;198: 111793.
[241] Lu Xiaoyang, Lin Peijie, Cheng Shuying, Lin Yaohai, Chen Zhicong, Wu Lijun, Zheng Qianying. Fault diagnosis for photovoltaic array based on convolutional neural network and electrical time series graph. Energy Convers Manage 2019; 196:950–65.
[242] Li Xiaoxia, Yang Qiang, Yan Wenjun, Chen Zhebo. Intelligent fault pattern recognition of aerial photovoltaic module images based on deep learning technique. J Syst Cybern Inf 2018;16(2):67–71.
[243] Li Xiaoxia, Li Wei, Yang Qiang, Yan Wenjun, Zomaya Albert Y. Building an online defect detection system for large-scale photovoltaic plants. In: Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation; 2019. p. 253–62.
[244] Li Xiaoxia, Yang Qiang, Lou Zhuo, Yan Wenjun. Deep learning based module defect analysis for large-scale photovoltaic farms. IEEE Trans Energy Convers 2018;34(1):520–9.
[245] Pierdicca R, Malinverni ES, Piccinini F, Paolanti M, Felicetti A, Zingaretti P. Deep convolutional neural network for automatic detection of damaged photovoltaic cells. Int Arch Photogramm, Remote Sens Spatial Informat Sci 2018;42(2).
[246] Alam Mohammed Khorshed, Khan Faisal, Johnson Jay, Flicker Jack. A comprehensive review of catastrophic faults in pv arrays: types, detection, and mitigation techniques. IEEE J Photovolt 2015;5(3):982–97.
[247] Lian Jijian, Zhang Yusheng, Ma Chao, Yang Yang, Chaima Evance. A review on recent sizing methodologies of hybrid renewable energy systems. Energy Convers Manage 2019;199:112027.
[248] Abba Lawan Bukar, Chee Wei Tan. A review on stand-alone photovoltaic wind energy system with fuel cell system optimization and energy management strategy. J Cleaner Prod, 2019.
[249] Rios Luis Miguel, Sahinidis Nikolaos V. Derivative-free optimization: a review of algorithms and comparison of software implementations. J Global Optim 2013;56 (3):1247–93.
[250] Arash Navaeefard SM, Tafreshi Moghaddas, Barzegari Mostafa, Shahrood Amir Jalali. Optimal sizing of distributed energy resources in microgrid considering wind energy uncertainty with respect to reliability. In: 2010 IEEE International Energy Conference. IEEE; 2010. p. 820–5.
[251] Tudu Bhimsen, Majumder Sibsankar, Mandal Kamal K, Chakraborty Niladri. Comparative performance study of genetic algorithm and particle swarm optimization applied on off-grid renewable hybrid energy system. In: International Conference on Swarm, Evolutionary, and Memetic Computing. Springer; 2011. p. 151–8.
[252] Hong Ying-Yi, Lian Ruo-Chen. Optimal sizing of hybrid wind/pv/diesel generation in a stand-alone power system using markov-based genetic algorithm. IEEE Trans Power Deliv 2012;27(2):640–7.
[253] Maleki Akbar, Pourfayaz Fathollah. Optimal sizing of autonomous hybrid photovoltaic/wind/battery power system with lpsp technology by using evolutionary algorithms. Sol Energy 2015;115:471–83.
[254] Ould Bilal B, Sambou V, Ndiaye PA, Kébé CMF, Ndongo M. Optimal design of a hybrid solar–wind-battery system using the minimization of the annualized cost system and the minimization of the loss of power supply probability (lpsp). Renewable Energy 2010;35(10):2388–90.
[255] Lan Hai, Wen Shuli, Hong Ying-Yi, David CYu, Zhang Lijun. Optimal sizing of hybrid pv/diesel/battery in ship power system. Appl Energy 2015;158:26–34.
[256] Nekkache Abdesslem, Bouzidi Belkacem, Kaabeche Abdelhamid, Bakelli Yahia. Hybrid pv-wind based water pumping system optimum sizing: a pso-llp-lpsp optimization and cost analysis. In: 2018 International Conference on Electrical Sciences and Technologies in Maghreb (CISTEM). IEEE; 2018. p. 1–6.
[257] Bouabdallah Ahmed, Bourguet Salvy, Olivier Jean-Christophe, Machmoum Mohamed. Optimal sizing of a stand-alone photovoltaic system. In: 2013 International Conference on Renewable Energy Research and Applications (ICRERA). IEEE; 2013. p. 543–8.
[258] Paliwal Priyanka, Patidar NP, Nema RK. Determination of reliability constrained optimal resource mix for an autonomous hybrid power system using particle swarm optimization. Renewable Energy 2014;63:194–204.
[259] Gharavi H, Ardehali MM, Ghanbari-Tichi S. Imperial competitive algorithm optimization of fuzzy multi-objective design of a hybrid green power system with considerations for economics, reliability, and environmental emissions. Renewable Energy 2015;78:427–37.
[260] Kanase-Patil AB, Saini RP, Sharma MP. Sizing of integrated renewable energy system based on load profiles and reliability index for the state of uttarakhand in india. Renewable Energy 2011;36(11):2809–21.
[261] Panayiotou Gregoris, Kalogirou Soteris, Tassou Savvas. Design and simulation of a pv and a pv–wind standalone energy system to power a household application. Renewable Energy 2012;37(1):355–63.
[262] Ahmadi Saeedeh, Abdi Shirzad. Application of the hybrid big bang–big crunch algorithm for optimal sizing of a stand-alone hybrid pv/wind/battery system. Sol Energy 2016;134:366–74.
[263] Ardakani Fatemeh Jahanbani, Riahy Gholamhossein, Abedi Mehrdad. Design of an optimum hybrid renewable energy system considering reliability indices. In: 2010 18th Iranian conference on electrical engineering. IEEE; 2010. p. 842–7.
[264] Kaabeche A, Belhamel M, Ibtiouen R. Sizing optimization of grid-independent hybrid photovoltaic/wind power generation system. Energy 2011;36(2):1214–22.
[265] Borhanazad Hanieh, Mekhilef Saad, Ganapathy Velappa Gounder, Modiri-Delshad Mostafa, Mirtaheri Ali. Optimization of micro-grid system using mopso. Renewable Energy 2014;71:295–306.
[266] Maleki Akbar, Askarzadeh Alireza. Artificial bee swarm optimization for optimum sizing of a stand-alone pv/wt/fc hybrid system considering lpsp concept. Sol Energy 2014;107:227–35.
[267] Zhang Debao, Liu Junwei, Jiao Shifei, Tian Hao, Lou Chengzhi, Zhou Zhihua, Zhang Ji, Wang Chendong, Zuo Jian. Research on the configuration and operation effect of the hybrid solar-wind-battery power generation system based on nsga-ii. Energy 2019;189:116121.
[268] Tegani I, Aboubou A, Becherif M, Ayad MY, Kraa O, Bahri M, Akhrif Ouassima. Optimal sizing study of hybrid wind/pv/diesel power generation unit using genetic algorithm. In: 4th International Conference on Power Engineering, Energy and Electrical Drives. IEEE; 2013. p. 134–40.
[269] Muhammad Yousif, Qian Ai, Yang Gao, Waqas Ahmad Wattoo, Ziqing Jiang, Ran Hao. An optimal dispatch strategy for distributed microgrids using pso. CSEE J Power Energy Syst, 2019.
[270] Katsigiannis Yiannis A, Georgilakis Pavlos S, Karapidakis Emmanuel S. Hybrid simulated annealing–tabu search method for optimal sizing of autonomous power systems with renewables. IEEE Trans Sustainable Energy 2012;3(3):330–8.
[271] Ramin Hosseinalizadeh, Hamed Shakouri, Mohsen Sadegh Amalnick, Peyman Taghipour. Economic sizing of a hybrid (pv–wt–fc) renewable energy system (hres) for stand-alone usages by an optimization-simulation model: Case study of Iran. Renewable Sustainable Energy Rev 2016;54:139–150.
[272] Baghaee Hamid Reza, Mirsalim Mojtaba, Gharehpetian Gevork B. Multi-objective optimal power management and sizing of a reliable wind/pv microgrid with hydrogen energy storage using mopso. J Intell Fuzzy Syst 2017;32(3):1753–73.
[273] Hlal Mohamed Izdin, Ramachandaramurthya Vigna K, Padmanaban Sanjeevikumar, Kaboli Hamid Reza, Pouryekta Aref, Abdullah Tuan, Rashid Tuan Ab. Nsga-ii and mopso based optimization for sizing of hybrid pv/wind/battery energy storage system. Int J Power Electron Drive Syst 2019;10(1): 463–78.
[274] Rahman Imran, Vasant Pandian M, Singh Balbir Singh Mahinder, Abdullah-Al-Wadud M. Novel metaheuristic optimization strategies for plug-in hybrid electric vehicles: A holistic review. Intell Decis Technol 2016;10(2):149–63.
[275] Ali Kaveh, Vahid Reza Mahdavi. Colliding bodies optimization: a novel meta-heuristic method. Comput Struct 2014;139:18–27.
[276] Glover F, Sörensen K. Metaheuristics. Scholarpedia 2015;10(4):6532. revision #149834.
[277] Ekren Orhan, Ekren Banu Y. Size optimization of a pv/wind hybrid energy conversion system with battery storage using simulated annealing. Appl Energy 2010;87(2):592–8.
[278] Kenneth Sorensen, Marc Sevaux, Fred Glover. A history of metaheuristics. arXiv preprint arXiv:1704.00853, 2017.
[279] Fetanat Abdolvahhab, Khorasaninejad Ehsan. Size optimization for hybrid photovoltaic–wind energy system using ant colony optimization for continuous domains based integer programming. Appl Soft Comput 2015;31:196–209.
[280] Suhane Payal, Rangnekar Saroj, Mittal Arvind, Khare Anula. Sizing and performance analysis of standalone wind-photovoltaic based hybrid energy system using ant colony optimisation. IET Renew Power Gener 2016;10(7): 964–72.
[281] Dong Weiqiang, Li Yanjun, Xiang Ji. Sizing of a stand-alone photovoltaic/wind energy system with hydrogen and battery storage based on improved ant colony

algorithm. In: 2016 Chinese Control and Decision Conference (CCDC). IEEE; 2016. p. 4461–6.
[282] Kalyani Manda, Satapathy S, Poornasatyanarayana B. Population based meta-heuristic techniques for solving optimization problems: A selective survey. Int J Emerg Technol Adv Eng 2012;2(11):206–211.
[283] Merei Ghada, Berger Cornelius, Sauer Dirk Uwe. Optimization of an off-grid hybrid pv–wind–diesel system with different battery technologies using genetic algorithm. Sol Energy 2013;97:460–73.
[284] Li Jiaming. Optimal sizing of grid-connected photovoltaic battery systems for residential houses in australia. Renewable Energy 2019;136:1245–54.
[285] Maleki Akbar, Ameri Mehran, Keynia Farshid. Scrutiny of multifarious particle swarm optimization for finding the optimal size of a pv/wind/battery hybrid system. Renewable Energy 2015;80:552–63.
[286] Bansal Ajay Kumar, Gupta RA, Kumar Rajesh. Optimization of hybrid pv/wind energy system using meta particle swarm optimization (mpso). In: India International Conference on Power Electronics 2010 (IICPE2010). IEEE; 2011. p. 1–7.
[287] Moghaddam Sasan, Bigdeli Mehdi, Moradlou Majid, Siano Pierluigi. Designing of stand-alone hybrid pv/wind/battery system using improved crow search algorithm considering reliability index. Int J Energy Environ Eng 2019;10(4): 429–49.
[288] Nadjemi O, Nacer T, Hamidat A, Salhi H. Optimal hybrid pv/wind energy system sizing: Application of cuckoo search algorithm for algerian dairy farms. Renew Sustain Energy Rev 2017;70:1352–65.
[289] Prakash Prem, Khatod Dheeraj K. Optimal sizing and siting techniques for distributed generation in distribution systems: A review. Renew Sustain Energy Rev 2016;57:111–30.
[290] Caasi John Kevin L, Aguirre Rodolfo A. Comparative analysis of the optimal siting and sizing on different solar distributed generation models through stochastic method. In: 2016 IEEE Innovative Smart Grid Technologies-Asia (ISGT-Asia), IEEE, 2016. p. 485–490.
[291] Dayapera Rei-Ann M, Aguirre Rodolfo A. Determination of penetration limit of solar distributed generation (dg) considering multiple bus integration. In: 2018 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC). IEEE; 2018. p. 508–13.
[292] Ahmadi Mikaeel, Lotfy Mohammed Elsayed, Shigenobu Ryuto, Yona Atsushi, Senjyu Tomonobu. Optimal sizing and placement of rooftop solar photovoltaic at kabul city real distribution network. IET Generat Transmiss Distrib 2017;12(2): 303–9.
[293] Raptis Dimitris A, Periandros Papamarkos S, Gkaidatzis Paschalis A, Bouhouras Aggelos S, Labridis Dimitris P. Optimal siting of bess in distribution networks under high pv penetration. In: 2018 53rd International Universities Power Engineering Conference (UPEC). IEEE; 2018. p. 1–6.
[294] Babacan Oytun, Torre William, Kleissl Jan. Siting and sizing of distributed energy storage to mitigate voltage impact by solar pv in distribution systems. Sol Energy 2017;146:199–208.
[295] Rodríguez-Gallegos Carlos D, Gandhi Oktoviano, Yang Dazhi, Alvarez-Alvarado Manuel S, Zhang Wenjie, Reindl Thomas, Panda Sanjib Kumar. A siting and sizing optimization approach for pv–battery–diesel hybrid systems. IEEE Trans Ind Appl 2017;54(3):2637–45.
[296] Niknam Taher, Taheri Seyed Iman, Aghaei Jamshid, Tabatabaei Sajad, Nayeripour Majid. A modified honey bee mating optimization algorithm for multiobjective placement of renewable energy resources. Appl Energy 2011;88 (12):4817–30.
[297] Abd El-salam, Mirna Fouad, Eman Beshr, Magdy B Eteiba. A new hybrid technique for minimizing power losses in a distribution system by optimal sizing and siting of distributed generators with network reconfiguration. Energies 2018; 11(12):3351.
[298] Mozafar Mostafa Rezaei, Moradi Mohammad H, Hadi Amini M. A simultaneous approach for optimal allocation of renewable energy sources and electric vehicle charging stations in smart grids based on improved ga-pso algorithm. Sustainable Cities Soc 2017;32:627–37.
[299] Moradi Mohammad H, Mohammad Abedini SM, Tousi Reza, Mahdi Hosseinian S. Optimal siting and sizing of renewable energy sources and charging stations simultaneously based on differential evolution algorithm. Int J Electrical Power Energy Syst 2015;73:1015–24.
[300] Karami Nabil, Moubayed Nazih, Outbib Rachid. General review and classification of different mppt techniques. Renew Sustain Energy Rev 2017;68:1–18.
[301] Salam Zainal, Ahmed Jubaer, Merugu Benny S. The application of soft computing methods for mppt of pv system: A technological and status review. Appl Energy 2013;107:135–48.
[302] Algazar Mohamed M, El-Halim Hamdy Abd, Salem Mohamed Ezzat El Kotb, et al. Maximum power point tracking using fuzzy logic control. Int J Electrical Power Energy Syst 2012;39(1):21–8.
[303] Farajdadian Shahriar, Hassan Hosseini SM. Optimization of fuzzy-based mppt controller via metaheuristic techniques for stand-alone pv systems. Int J Hydrogen Energy 2019;44(47):25457–72.
[304] Li Xingshuo, Wen Huiqing, Yihua Hu, Jiang Lin. A novel beta parameter based fuzzy-logic controller for photovoltaic mppt application. Renewable Energy 2019; 130:416–27.
[305] Liu Yi-Hua, Liu Chun-Liang, Huang Jia-Wei, Chen Jing-Hsiau. Neural-network-based maximum power point tracking methods for photovoltaic systems operating under fast changing environments. Sol Energy 2013;89:42–53.
[306] Mahmoud Nour Ali. Improved design of artificial neural network for mppt of grid-connected pv systems. In: 2018 Twentieth International Middle East Power Systems Conference (MEPCON), EEE, 2018. p. 97–102.
[307] Messalti Sabir, Harrag Abdelghani, Loukriz Abdelhamid. A new variable step size neural networks mppt controller: Review, simulation and hardware implementation. Renew Sustain Energy Rev 2017;68:221–33.
[308] Chao Kuei-Hsiang, Lin Yu-Sheng, Lai Uei-Dar. Improved particle swarm optimization for maximum power point tracking in photovoltaic module arrays. Appl Energy 2015;158:609–18.
[309] Sudhakar Babu T, Rajasekar N, Sangeetha K. Modified particle swarm optimization technique based maximum power point tracking for uniform and under partial shading condition. Appl Soft Comput 2015;34:613–24.
[310] Eltamaly Ali M, Farh Hassan MH, Al Saud Mamdooh S. Impact of pso reinitialization on the accuracy of dynamic global maximum power detection of variant partially shaded pv systems. Sustainability 2019;11(7):2091.
[311] Besheer AH, Adly M. Ant colony system based pi maximum power point tracking for stand alone photovoltaic system. In: 2012 IEEE International Conference on Industrial Technology. IEEE; 2012. p. 693–8.
[312] Sundareswaran Kinattingal, Vigneshkumar Vethanayagam, Sankar Peddapati, Simon Sishaj P, Srinivasa Rao Nayak P, Palani Sankaran. Development of an improved p&o algorithm assisted through a colony of foraging ants for mppt in pv system. IEEE Trans Industr Inf 2015;12(1):187–200.
[313] Titri Sabrina, Larbes Cherif, Toumi Kamal Youcef, Benatchba Karima. A new mppt controller based on the ant colony optimization algorithm for photovoltaic systems under partial shading conditions. Appl Soft Comput 2017;58:465–79.
[314] Salih Hasan Wahhab, Wang Shaorong, Farhan Bashar Sakeen. A novel ga-pi optimized controller for mppt based pv in a hybrid pv-diesel power system. In: 2015 5th International Conference on Electric Utility Deregulation and Restructuring and Power Technologies (DRPT). IEEE; 2015. p. 1288–93.
[315] Abderraouf Messai, Adel Mellit, Guessoum A, Kalogirou SA. Maximum power point tracking using a ga optimized fuzzy logic controller and its fpga implementation. Sol Energy 2011;85(2):265–277.
[316] Prasad Lal Bahadur, Sahu Suneet, Gupta Monika, Srivastava Rishabh, Mozhui Lichamo, Asthana Dhawal N. An improved method for mppt using ann and ga with maximum power comparison through perturb & observe technique. In: 2016 IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics Engineering (UPCON). IEEE; 2016. p. 206–11.
[317] Tey Kok Soon, Mekhilef Saad, Seyedmahmoudian Mehdi, Horan Ben, Oo Amanullah Than, Stojcevski Alex. Improved differential evolution-based mppt algorithm using sepic for pv systems under partial shading conditions and load variation. IEEE Trans Industr Inf 2018;14(10):4322–33.
[318] Sundareswaran Kinattingal, Peddapati Sankar, Palani Sankaran. Mppt of pv systems under partial shaded conditions through a colony of flashing fireflies. IEEE Trans Energy Convers 2014;29(2):463–72.
[319] Ahmed Jubaer, Salam Zainal. A maximum power point tracking (mppt) for pv system using cuckoo search with partial shading capability. Appl Energy 2014; 119:118–30.
[320] Hassan Muhammed A, Khalil A, Kaseb S, Kassem MA. Exploring the potential of tree-based ensemble methods in solar radiation modeling. Appl Energy 2017;203: 897–916.
[321] Ibrahim Anwar Ibrahim, Tamer Khatib. A novel hybrid model for hourly global solar radiation prediction using random forests technique and firefly algorithm. Energy Convers Manage 2017;138:413–425.
[322] Linares-Rodriguez Alvaro, Quesada-Ruiz Samuel, Pozo-Vazquez David, Tovar-Pescador Joaquin. An evolutionary artificial neural network ensemble model for estimating hourly direct normal irradiances from meteosat imagery. Energy 2015; 91:264–73.
[323] Eissa Yehia, Marpu Prashanth R, Gherboudj Imen, Ghedira Hosni, Ouarda Taha BMJ, Chiesa Matteo. Artificial neural network based model for retrieval of the direct normal, diffuse horizontal and global horizontal irradiances using seviri images. Sol Energy 2013;89:1–16.
[324] Khosravi A, Koury RNN, Machado L, Pabon JJG. Prediction of hourly solar radiation in abu musa island using machine learning algorithms. J Clean Prod 2018;176:63–75.
[325] Lou Siwei, Li Danny HW, Lam Joseph C, Chan Wilco WH. Prediction of diffuse solar irradiance using machine learning and multivariable regression. Appl Energy 2016;181:367–74.
[326] Jallal Mohammed Ali, Chabaa Samira, Zeroual Abdelouhab. A new artificial multi-neural approach to estimate the hourly global solar radiation in a semi-arid climate site. Theoret Appl Climatol 2020;139(3–4):1261–76.
[327] Aler Ricardo, Galván Inés M, Ruiz-Arias Jose A, Gueymard Christian A. Improving the separation of direct and diffuse solar radiation components using machine learning by gradient boosting. Sol Energy 2017;150:558–69.
[328] Ramli Makbul AM, Ssennoga Twaha, Al-Turki Yusuf A. Investigating the performance of support vector machine and artificial neural networks in predicting solar radiation on a tilted surface: Saudi arabia case study. Energy Convers Manage 2015;105:442–452.
[329] Khatib Tamer, Elmenreich Wilfried. A model for hourly solar radiation data generation from daily solar radiation data using a generalized regression artificial neural network. Int J Photoenergy 2015;2015.
[330] Zhang Wenqi, Kleiber William, Florita Anthony R, Hodge Bri-Mathias, Mather Barry. A stochastic downscaling approach for generating high-frequency solar irradiance scenarios. Sol Energy 2018;176:370–9.
[331] Frimane Àzeddine, Soubdhan Ted, Bright Jamie M, Aggour Mohammed. Nonparametric bayesian-based recognition of solar irradiance conditions: Application to the generation of high temporal resolution synthetic solar irradiance data. Sol Energy 2019;182:462–79.

[332] Widén Joakim, Munkhammar Joakim. Spatio-temporal downscaling of hourly solar irradiance data using gaussian copulas. In: 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC). IEEE; 2019. p. 3172–8.
[333] Assouline Dan, Mohajeri Nahid, Scartezzini Jean-Louis. Quantifying rooftop photovoltaic solar energy potential: A machine learning approach. Sol Energy 2017;141:278–96.
[334] Mohajeri Nahid, Assouline Dan, Guiboud Berenice, Bill Andreas, Gudmundsson Agust, Scartezzini Jean-Louis. A city-scale roof shape classification using machine learning for solar energy applications. Renewable Energy 2018; 121:81–93.
[335] Assouline Dan, Mohajeri Nahid, Scartezzini Jean-Louis. Large-scale rooftop solar photovoltaic technical potential estimation using random forests. Appl Energy 2018;217:189–211.
[336] Yang Dazhi. Solardata: An r package for easy access of publicly available solar datasets. Sol Energy 2018;171:A3–12.
[337] Sheffield Solar. Microgen database. Sheffield Solar-University of Sheffield, [Online]. Available: http://www.microgen-database.org.uk, 2016.
[338] Bright Jamie M, Killinger Sven, Engerer Nicholas A. Data article: Distributed pv power data for three cities in australia. J Renewable Sustainable Energy 2019;11 (3):035504.
[339] Pedro Hugo TC, Larson David P, Coimbra Carlos FM. A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods. J Renewable Sustainable Energy 2019;11(3):036102.
[340] Yu Manajit Sengupta, Xie Anthony Lopez, Habte Aron, Maclaurin Galen, Shelby James. The national solar radiation data base (nsrdb). Renew Sustain Energy Rev 2018;89:51–60.
[341] Stoffel T, Andreas A. Nrel solar radiation research laboratory (srrl): Baseline measurement system (bms); golden, colorado (data). Technical report, National Renewable Energy Lab. (NREL), Golden, CO (United States), 1981.
[342] Augustine John A, DeLuisi John J, Long Charles N. Surfrad–a national surface radiation budget network for atmospheric research. Bull Am Meteorol Soc 2000; 81(10):2341–58.
[343] Bradbury Kyle, Saboo Raghav, Johnson Timothy L, Malof Jordan M, Devarajan Arjun, Zhang Wuming, et al. Distributed solar photovoltaic array location and extent dataset for remote sensing object identification. Sci Data 2016;3(1):1–9.